

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Classification et admissibilité

Derouf, V.

Award date:
1982

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

FACULTES UNIVERSITAIRES N.D. DE LA PAIX

NAMUR

FACULTE DES SCIENCES

1981-1982

CLASSIFICATION
ET
ADMISSIBILITE

Mémoire présentée pour l'obtention
du grade de Licence en Sciences
mathématiques
par

promoteur :

R. J.P. Rasson

Véronique DEROUF

que le Professeur Ramon et
Monsieur Houdy trouvent ici
l'expression de mes plus vifs
remerciements.

Ils tiennent également à remercier
le Professeur Arday.

Étant donné, le nombre important de méthodes de classification, il est avantageux de connaître les tendances et propriétés de chacune d'entre elles pour guider l'utilisateur dans le choix d'une procédure adaptée à son problème. Après avoir introduit la classification automatique, nous allons dans un premier chapitre, de'exire quelques méthodes de classification. Ensuite, nous allons étudier les conditions d'administrabilité de ces procédures d'un point de vue théorique dans le deuxième chapitre avec les neuf conditions d'administrabilité proposées par Fieser et Van Nieu (1971), et d'un point de vue pratique dans le chapitre trois. Dans un dernier chapitre, nous présenterons d'autres conditions d'administrabilité.

Table des matières

Ch 0 Introduction

0.1	la classification automatique	1
0.2	un peu d'histoire	2
0.3	quelques remarques	4
0.4	Formulation du problème et définition	6
	de base	

Ch 1 Méthodes de classification

§ 1.	les méthodes hiérarchiques	
1.	les méthodes - définitions	12
2.	les indices d'aggrégation	18
3.	la formule de réécriture	19

§ 2. les méthodes non hiérarchiques

	optimisant un critère	
A.	Méthodes de l'axe de la somme	23
	des carrés et du K. means	
B.	la méthode des aires	30

Ch 2. Conditions d'admissibilité : point de vue théorique

§ 1 Présentation des conditions d'admissibilité 38

§ 2 Vérification des conditions

A. Méthodes hiérarchiques 54

B. Méthodes de l'erreur de la somme 88

C. Méthode des surfaces des carrés et du K-means 100

Ch 3. Conditions d'admissibilité point de vue pratique

1. Présentation des ensembles testés 108

2. Les méthodes et les données 111

3. Tendances des procédures 117

4. A propos des méthodes hiérarchiques 123

5. A propos des méthodes optimisant un 125

critère

Ch 4. Autres conditions d'admissibilité

§ 1 Présentation des conditions concernant l'aspect

littéraire

§ 2 Présentation des conditions concernant l'aspect

musique

CONCLUSION

ANNEXES

INTRODUCTION

0.1 La classification automatique

La nature offre un grand nombre de populations qu'il est souhaitable de répartir en catégories; chaque discipline scientifique adolite des classifications; il peut n'agir, en médecine par exemple, de découvrir les principaux regroupements de malades ayant le même comportement vis à vis de certains paramètres retenus pour les caractériser. En botanique, il n'agit de mettre, par exemple, en évidence des sous-espèces d'une même variété a partir d'un tableau de données où les plantes sont caractérisées par un certain nombre de mesures.

La classification automatique est l'ensemble des procédés qui permettent de regrouper des objets donnés en classes (en anglais: "clusters") telles que les éléments d'une même classe soient similaires, semblables par rapport aux paramètres les caractérisant. On entend également par classification, les classes

produites par une certaine procédure. Les techniques qui
généralisent des classifications sont connues de différentes
manières, comme taxonomie murmurée, typologie,
"cluster analysis", "Q-analysis", ... Cette variété de
noms est due à l'importance de ces méthodes dans
des champs aussi divers que la psychologie, la
zoologie, la sociologie, l'économie, ...

0.2 Un peu d'histoire

La nécessité de classer, de regrouper ou de
reconstruire des objets sur lesquels on possède des
informations existe depuis très longtemps. Déjà
les Grecs et les Romains développaient différents typ-
ologies humaines basées sur les caractéristiques
physiques. La classification a joué un rôle central
dans de nombreux domaines; par exemple, la
telle périodique en chimie produite en grande
partie par Mendeleev en 1869, a eu une
profonde influence sur la compréhension de la
structure atomique.

les méthodes étaient basées sur la façon dont on distinguait le semblable du dissimilaire. E' est toujours à la fin du 18^{ème} siècle que le biologiste Adamson donnait le principe d'un des premiers algorithmes de classification automatique. "Je me contentais de rapprocher les objets suivant le plus grand nombre de degrés de leurs rapports et de leurs ressemblances. Les objets ainsi réunis formaient plusieurs petites familles que je réunissais encore ensemble afin d'en faire un tout dont les parties soient unies et liées intimement..." Jusqu'au 20^{ème} siècle, la classification était plus un art qu'une méthode scientifique. E' est à partir des idées d'Adamson que Akal & Ansell (1963) introduisent les types idéologiques [cht] en biologie pour lesquels la classification est basée sur toute les caractéristiques des individus. E' est depuis l'apparition de l'ordinateur (1960) que le nombre de techniques de classification a considérablement augmenté. Beaucoup de ces méthodes reposent sur des fondements essentiellement géométriques ou algébriques et ne font pas intervenir de modèle

Il est intéressant de distinguer "classification" et "direction". En direction, les données ne forment qu'un seul groupe et on décide de le décaper en différents acteurs ayant chacun des particularités propres. Des méthodes de classification sont quelque-fois utilisées dans un tel contexte mais le but vise

0.3 Quelques remarques.

probabiliste juge par certains comme beaucoup trop contraignant. Un antagonisme n'est développé entre les porteurs de la stratégie inférentielle et ceux de l'analyse de données sans modèle probabiliste. Les deux tendances extrêmes devraient s'équilibrer avec le temps car la stratégie a besoin de techniques pour mener de description des données s'appuyant sur des possibilités nouvelles de l'ordinateur. Elle a aussi besoin de modèles inférentiels adéquats pour la stratégie non paramétrique pour valider les résultats obtenus, comparer les méthodes, étudier les comportements asymptotiques...

par la classification n'est pas d'importance majeure. On peut également remarquer que plusieurs

définitions de classes ("clusters") sont proposées. En voici quelques unes :

- Une classe est un ensemble d'objets se ressemblant le plus possible et étant le plus différent possible des objets des autres classes

- Une classe est un ensemble d'objets qui peuvent être traités de manière équivalente (Wallerstein, Bollen)

- Une classe est une collection de points tels que la distance entre deux points quelconques de la

classe soit plus petite que la distance entre un

point appartenant à la classe et un point n'y

appartenant pas (Gengst)

- Les classes doivent être aussi homogènes que

possible et aussi séparées que possible.

Il n'y a pas de convention universelle pour définir

une classe. Il est généralement vrai que plus d'une

solution soient correctes. Il s'agit d'une situation pour laquelle

la signification d'un tel terme est le jugement de

l'utilisateur. Tout dépend de ce qu'il veut faire !

0.4 Formulation du problème et définition de base.

Considérons M un ensemble de m individus ou objets, x_1, x_2, \dots, x_n . Pour chacun de ces individus, on mesure p caractéristiques. Les données peuvent alors se mettre sous la forme d'un tableau

$$X = (x_{ij}) \quad i = 1, \dots, m \quad j = 1, \dots, p$$

où x_{ij} est la valeur de la $j^{\text{ème}}$ caractéristique pour le $i^{\text{ème}}$ individu.

Dans la suite on considèrera l'individu i et le vecteur le caractérisant

$$\begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$$

Toute application qui permet d'exprimer numériquement un lien existant entre les individus ou entre les variables est appelée une "mesure de ressemblance". On présente ici quelques définitions de base qui serviront dans la suite :

Définition d'un indice de similitude

Un indice de similitude sur un ensemble M est une application α qui vérifie les trois propriétés suivantes :

1. α est une application $M \times M \rightarrow \mathbb{R}^+$
2. α est symétrique : $\forall (x_i, x_j) \in M \times M$
 $\alpha(x_i, x_j) = \alpha(x_j, x_i)$
3. $\forall x_i, x_j \in M$ avec $i \neq j$
 $\alpha(x_i, x_i) = \alpha(x_j, x_j) > \alpha(x_i, x_j)$

Définition d'un indice de dissimilitude

Un indice de dissimilitude est une application d qui satisfait aux conditions données dans la définition d'une mesure de dissimilitude, sauf la troisième qui est remplacée par :

$\forall x_i \in M \quad d(x_i, x_i) = 0.$

Définition d'une distance et d'une ultramétrique.

Une distance est un indice de dissimilarité qui vérifie en plus les deux propriétés suivantes :

1. $d(x, y) = 0 \Leftrightarrow x = y$ $x, y \in M$
2. $d(x, y) \leq d(x, z) + d(z, y)$ $\forall x, y, z \in M$

(inégalité triangulaire)

Une distance ultramétrique est une distance qui au lieu de vérifier l'inégalité triangulaire vérifie

$$d(x, y) \leq \max \{ d(x, z), d(z, y) \} \quad \forall x, y, z \in M$$

Exemple

Prenons $M = \{ (0, 0), (2, 0), (1, 4) \}$
la distance euclidienne est une ultramétrique
sur M .

METHODES DE CLASSIFICATION

Un dictionnaire, en général, deux grands concepts de classes :

- les classes monothétiques
- les classes polythétiques

On dit qu'une classe est monothétique si elle est caractérisée par une propriété, ce qui signifie que tous les éléments de la classe prennent la même valeur pour cette propriété, ou le contraire.

Il intervient des procédés fournissant des classes monothétiques est quand la classification

a leur dérivé, où il n'agit d'affecter à une classe un élément qui se présente positivement à une classification car on dispose alors de deux d'identification. On peut remarquer que les

classes de mots dans le dictionnaire sont

monothétiques. Mais une telle définition de classe

est trop rigide lorsqu'on se réfère aux classifications

naturelles dans les sciences humaines. Une
classe peut être lue naturelle sans qu'il soit
possible de définir un caractère commun à tous
les éléments la composant.

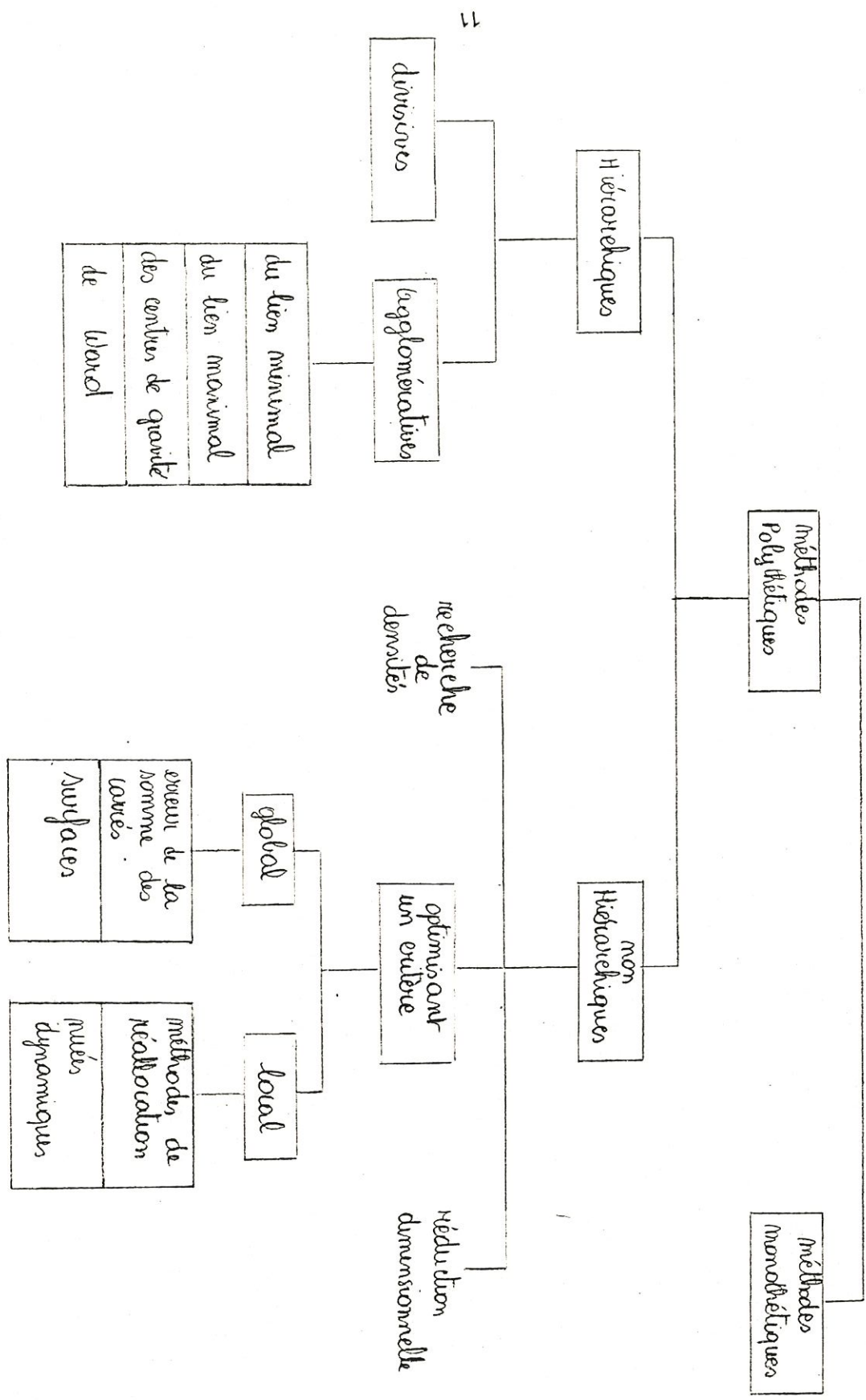
On opère à cette notion de classe monothétique
elle de classe polythétique où les éléments n
ressemblent relativement à l'ensemble des caractères.
La ressemblance des individus d'une classe est
considérée globalement, en général il n'existe pas
de caractères particuliers prenant une même valeur
pour tous les individus.

Parmi les techniques fournissant des classes poly-
thétiques, on peut distinguer deux grandes catégories.

- les méthodes hiérarchiques
- comprenant des méthodes agglomératives et
divisives

- les méthodes non hiérarchiques
- avec principalement des méthodes optimisant
localement ou globalement un critère, des
méthodes de recherche du meilleur ou de
réduction dimensionnelle.

Tableau des méthodes étudiées



§ 1 Les méthodes hiérarchiques

Généralement les méthodes hiérarchiques sont subdivisées en méthodes agglomératives et méthodes divisives.

Un algorithme est agglomératif si la construction des classes se fait par regroupement successif des constituants. Il construit des partitions de moins en moins fines.

Un algorithme est divisif s'il fonctionne successivement les classes. Il permet de construire des partitions de plus en plus fines.

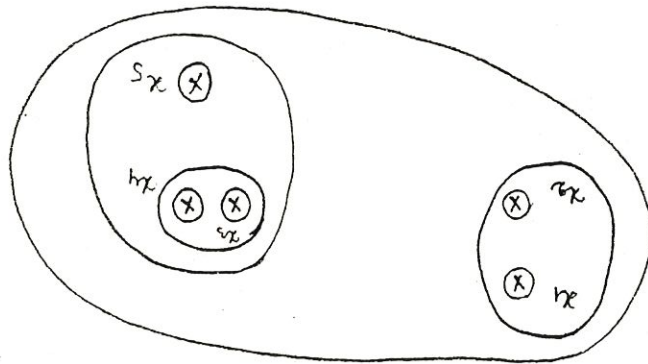
Dans le cadre de ce travail, nous nous limiterons à l'étude des méthodes agglomératives.

1. Les méthodes - définition

On cherche à regrouper des individus x_1, x_2, \dots, x_n de M par un ensemble de règles hiérarchiquement imbriquées.

Dans les procédures hiérarchiques agglomératives, on commence par considérer les m objets comme les m premières classes. On regroupe ensuite les deux plus proches afin d'obtenir $(m-1)$ classes. Le procédé est répété jusqu'à ce que tous les objets soient dans une même classe.

Pas exemple, si on considère l'ensemble M constitué des cinq éléments ci-après



à chaque étape, on regroupe les individus les plus proches

l'un ensemble de parties proches est donc

$$H = \left\{ \{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_1, x_2\}, \{x_3, x_4\}, \{x_4, x_5\}, \{x_1, x_2, x_3, x_4, x_5\} \right\}$$

H est appelée une hiérarchie.

de manière plus précise, on peut définir une hiérarchie de la façon suivante :

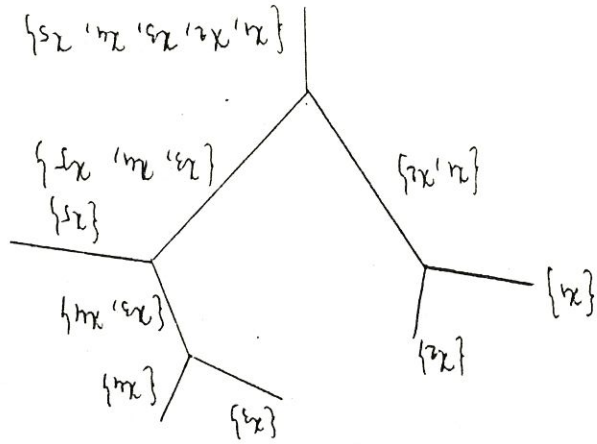
Définition d'une hiérarchie

Soit M un ensemble fini, H un ensemble de parties de M . H est une hiérarchie sur M si

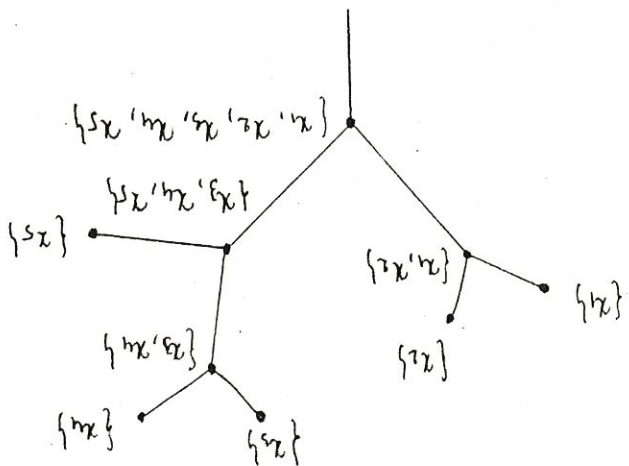
1. $M \in H$
2. $\forall x_i \in M, \{x_i\} \in H$
3. $\forall h, h' \in M$ on a : $h \cap h' \neq \emptyset \Rightarrow h \subset h' \text{ ou } h' \subset h$.

L'ordre de classification permet de visualiser les liens obtenus par une méthode hiérarchique. Il peut être dessiné de plusieurs façons :

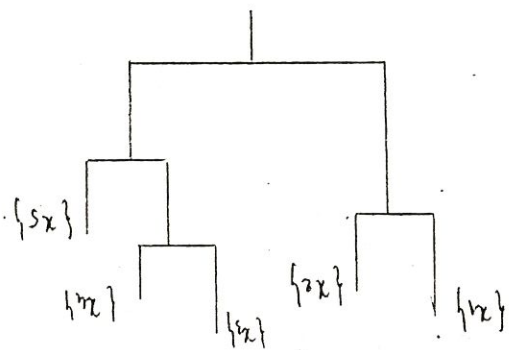
- Chaque élément peut être représenté par une branche



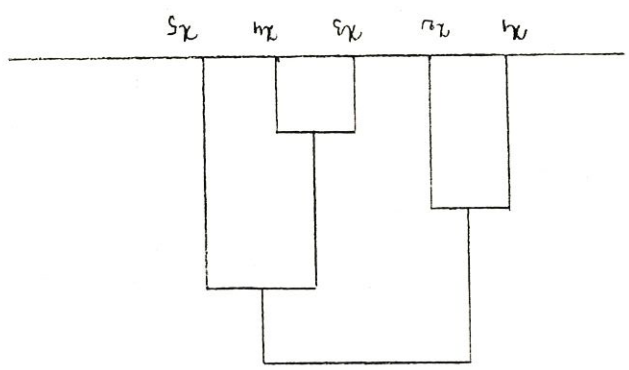
- Chaque élément peut être représenté par un nœud



- Les branches peuvent être verticales et les nœuds représentés par des traits horizontaux.



- D'ailleurs, on utilisera la représentation ci-dessous



Chaque palier de la hiérarchie nous tend un groupe de points. La hauteur du palier est une mesure du degré d'aggrégation du groupe qu'il nous tend. On voit aisément que les propriétés de la définition de hiérarchie sont satisfaites.

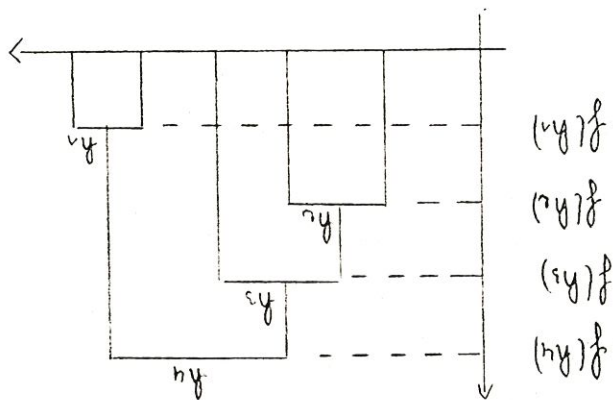
Comme pour chaque palier, on peut faire correspondre à sa hauteur une mesure d'aggrégation, on est amené à définir une hiérarchie indexée.

Définition d'une hiérarchie indexée

Une hiérarchie indexée est un couple (H, f) où H est une hiérarchie et f une application de H dans \mathbb{R}^+ telle que

1. $f(h) = 0 \Leftrightarrow h$ ne contient qu'un seul élément
2. $\forall h, h' \in H : h \subset h' \Rightarrow f(h) < f(h')$

Cette hiérarchie est bien indexée



Il peut se produire que l'indice amène à certains recherches donne lieu à l'existence de paliers
 pour h' tel que

$$h < h' \Rightarrow f(h) \leq f(h')$$

on dit alors que la fonction est indexée au

sens large

$$h < h' \Rightarrow f(h) > f(h')$$

on dit alors qu'il y a inversion.

La construction d'une hiérarchie nécessite la connaissance d'une mesure de ressemblance entre les groupes d'individus. Cette mesure est appelée indice d'aggrégation.

On appelle indice d'aggrégation entre groupes d'individus, une application $S: \mathcal{G}(M) \times \mathcal{G}(M) \rightarrow \mathbb{R}^+$

telles que : $\forall h_i, h_j \in H \quad S(h_i, h_j) = S(h_j, h_i)$

Les indices d'aggrégation sont généralement construits

à partir de l'indice de dissimilarité entre individus

de M , choisi par l'utilisateur.

Notons, que dans la suite nous la désignerons

successivement a été utilisée.

2. Les indices d'agrégation

l'indice d'agrégation du lien minimal (Gardner et Sibson)

$$\delta(x_i, y_j) = \min_{\substack{x_i \in R_i \\ y_j \in R_j}} d(x_i, y_j)$$

l'indice d'agrégation du lien maximal (Sorensen)

$$\delta(x_i, y_j) = \max_{\substack{x_i \in R_i \\ y_j \in R_j}} d(x_i, y_j)$$

l'indice d'agrégation des centres de gravité

$$\delta(x_i, y_j) = d(G(x_i), G(y_j))$$

avec $G(x_i)$ le centre de gravité de R_i

$$G(x_i) = \frac{\sum_{x_j \in R_i} x_j}{n_i}$$

l'indice d'agrégation de l'augmentation d'entité ou de poids

Cette formule permet une grande économie de temps machine et de place mémoire. Elle généralise la formule énoncée de Rame et Williams (1967) qui ne considère par les termes a_1, a_5, a_6 .
Regardons la valeur des coefficients pour les indices d'aggrégation étudiés :

$$\delta(h, h_i \cup h_j) = a_1 \delta(h, h_i) + a_2 \delta(h, h_j) + a_3 \delta(h_i, h_j) + a_4 f(h) + a_5 f(h_i) + a_6 f(h_j) + a_7 |\delta(h, h_i) - \delta(h, h_j)|$$

Les mesures de distances entre groupes d'individus satisfont à une formule de réécriture pour la distance entre un groupe h et un groupe formé par la fusion de deux classes h_i et h_j .

3. La formule de réécriture générale (Yambaru 1978)

avec $p(h_i)$ le poids de h_i
on peut écrire souvent $p(h_i) = m_i = \text{card}(h_i)$

$$\delta(h_i, h_j) = \frac{p(h_i) \cdot p(h_j)}{p(h_i) + p(h_j)} d(G(h_i), G(h_j))$$

R' indice du lien minimal

$$\min_{x \in R} d(x, y) = \frac{1}{2} \min d(x, y_i) + \frac{1}{2} \min d(x, y_j) - \frac{1}{2} | \min d(x, y_i) - \min d(x, y_j) |$$

$(y_i \in R_i, y_j \in R_j)$

on obtient directement $a_1 = a_2 = \frac{1}{2}$

$$a_T = -\frac{1}{2}$$

R' indice du lien maximal

$$\max_{x \in R} d(x, y) = \frac{1}{2} \max_{x \in R} d(x, y_i) + \frac{1}{2} \max_{x \in R} d(x, y_j) - \frac{1}{2} | \max_{x \in R} d(x, y_i) - \max_{x \in R} d(x, y_j) |$$

$(y_i \in R_i, y_j \in R_j)$

on obtient donc $a_1 = a_2 = \frac{1}{2}$

$$a_T = \frac{1}{2}$$

R' indice des centres de gravité

On utilise la formule suivante déduite du théorème d'Huygens [annexe 1]

$$a_3 = \frac{-m_k + m_i + m_j}{m_k}$$

$$a_1 = \frac{m_k + m_i + m_j}{m_k + m_i}$$

Les coefficients sont donc

$$a_2 = \frac{m_k + m_i + m_j}{m_k + m_j}$$

par (1) nous obtenons

$$\delta(x_i, x_i \cup x_j) = \frac{m_k + m_i + m_j}{m_k} [m_i d(G(x_i), G(x_j)) + m_j d(G(x_i), G(x_j)) - \frac{m_i m_j}{m_i + m_j} d(G(x_i), G(x_j))]$$

$$\delta(x_i, x_i \cup x_j) = \frac{m_k (m_i + m_j)}{m_k (m_i + m_j)} d(G(x_i), G(x_i \cup x_j))$$

l'indice de Ward

$$a_3 = \frac{-m_i m_j}{m_i + m_j}$$

$$a_1 = \frac{m_i + m_j}{m_i}$$

Les coefficients sont donc

$$a_2 = \frac{m_j}{m_i + m_j}$$

(1)

$$(m_i + m_j) d(G(x_i \cup x_j), G(x_i)) = m_i d(G(x_i), G(x_i)) + m_j d(G(x_i), G(x_j)) - \frac{m_i m_j}{m_i + m_j} d(G(x_i), G(x_j))$$

Requisitos

a_1	lien minimal	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{m_i + m_j}{m_i + m_j + m_k}$	Wald
a_2	lien maximal	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{m_j}{m_i + m_j}$	
a_3		0	0	0	$\frac{-m_i m_j}{m_i + m_j}$	
a_4		$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0

Pour les 4 procédures : $a_4 = a_5 = a_6 = 0$

§ 2. Méthodes non hiérarchiques, optimisant un critère

† Méthodes de l'erreur de la somme des carrés et du K. means

Supposons que l'on ait un ensemble de m individus x_1, x_2, \dots, x_n notés M_i , que l'on désire partitionner exactement en k classes C_1, C_2, \dots, C_k . Une façon de décrire le problème correctement est de définir une fonction qui attribue à chaque individu une qualité de chaque partitionnement des données. Le problème revient alors à trouver la partition qui optimise le critère. Un des critères les plus souvent utilisés est celui de "l'erreur de la somme des carrés" (sum-of-squared error). Regardons en quoi il consiste :

Appelons m_i le nombre d'individus dans la classe C_i . m_i la moyenne de cette classe

$$m_i = \frac{1}{m_i} \sum_{x_j \in C_i} x_j$$

L'erreur de la somme des carrés est définie par

$$\mathcal{L}(g, a_1, \dots, a_k) = \sum_{j=1}^k \left(\sum_{x_i \in g} \|x_i - m_g\|^2 \right)$$

Il est relativement aisé d'intépréter le critère à

minimiser. Pour une classe donnée, soit g , la

moyenne m_g est le meilleur représentant de la

classe en ce sens qu'il minimise la somme des

carrés des écarts entre les éléments de la classe

et le centre. $\mathcal{L}(g, a_1, \dots, a_k)$ mesure l'erreur totale

encourue quand on représente les m individus

x_1, x_2, \dots, x_m par les k centres des classes, m_1, m_2, \dots, m_k .

La position optimale est celle qui minimise le

critère.

Plusieurs approches sont possibles.

La méthode du minimum des carrés est fixée

cette procédure consiste à rechercher toutes les positions
possibles des m données en k classes, de calculer pour
chaque d'elles le valeur du critère \mathcal{L} .
La classification choisie est celle qui minimise le

l'idée de base est de partir d'une position initiale et de transférer les individus d'une classe à une autre si ils permettent la diminution de la valeur du critère. On continue la procédure jusqu'à ce qu'il n'y ait aucun point ne change de classe. Cette approche due à Régnier garantit un minimum

b) Méthode du "Hill Climbing" (K. Muramatsu)

l'optimisation locale
ici nous arrivons à l'étude d'une méthode
le nombre prend les valeurs importantes
 $S(15, 3) = 2.375.101$
 $S(59, 2) \approx 10^{18}$
 $S(100, 5) \approx 10^{68}$

[annexe 2]
est le nombre de strings qui vont
le nombre de positions de m éléments en k classes
elle n'est pas utilisable en pratique. En effet,
pour les données dans 10^m mais malheureusement
$$S(m, k) = \frac{1}{k} \sum_{i=1}^k (-1)^{k-i} \binom{k}{i} i^m$$

le critère devient alors

$$\mathcal{L}(a, \dots, a, \dots, c^*, y^*, \dots, y^*, b) = \sum_{j=1}^p \left(\sum_{i \in y_j} \|x_i - m_j\|^2 \right) + \sum_{i \in c^*} \|x_i - m_{c^*}\|^2 + \sum_{i \in y^*} \|x_i - m_{y^*}\|^2$$

$$m_{c^*}^* = m_c - \frac{m_c - 1}{p + m_c}$$

$$m_{y^*}^* = m_y + \frac{p - m_y}{m_y + 1}$$

Supposons qu'un point p actuellement dans la classe c passe dans la classe y .
Les moyennes des deux classes deviennent :

$$m_y = \frac{1}{n_y} \sum_{i \in y} x_i$$

$$\mathcal{L}(a, a, \dots, a, \dots, b) = \sum_{j=1}^p \left(\sum_{i \in y_j} \|x_i - m_j\|^2 \right)$$

local du critère.
Regardons de plus près ce que cela signifie. Reprenons que le critère est défini de la façon suivante :

avec $\mathcal{C}^* = \mathcal{C} \setminus \{p\}$
 $\mathcal{Q}^* = \mathcal{Q} \cup \{p\}$

$$\text{or } \sum_{x \in \mathcal{C}^*} \|x - m_{\mathcal{C}^*}\|^2 = \sum_{x \in \mathcal{C}} \|x - m_{\mathcal{C}}\|^2 - \|p - m_{\mathcal{C}}\|^2$$

on remplace $m_{\mathcal{C}^*}$ par $m_{\mathcal{C}} - \frac{p - m_{\mathcal{C}}}{n-1}$

$$= \sum_{x \in \mathcal{C}} \|x - m_{\mathcal{C}}\|^2 + \frac{n}{(n-1)^2} \|p - m_{\mathcal{C}}\|^2$$

$$+ n \sum_{x \in \mathcal{C}} \langle x - m_{\mathcal{C}}, p - m_{\mathcal{C}} \rangle - \|p - m_{\mathcal{C}}\|^2$$

$$\text{comme } \sum_{x \in \mathcal{C}} (x - m_{\mathcal{C}}) = 0$$

$$= \sum_{x \in \mathcal{C}} \|x - m_{\mathcal{C}}\|^2 - \frac{n}{n-1} \|p - m_{\mathcal{C}}\|^2$$

de manière analogue on trouve

$$\sum_{x \in \mathcal{Q}^*} \|x - m_{\mathcal{Q}^*}\|^2 = \sum_{x \in \mathcal{Q}} \|x - m_{\mathcal{Q}}\|^2$$

$$+ \frac{n}{n-1} \|p - m_{\mathcal{Q}}\|^2$$

le qui implique que

$$\mathcal{L}(a_1, \dots, a_r, q_1^*, \dots, q_r^*, q) = \mathcal{L}(a_1, a_2, \dots, a_r, q) + \frac{m}{m+1} \|p - mq\|^2$$

$$- \frac{m}{m+1} \|p - mq\|^2$$

le transfert de p de la classe c dans la classe q est avantageux n'il permet de faire décroître la valeur du critère. c'est à dire si

$$\frac{m}{m+1} \|p - mq\|^2 < \frac{m}{m+1} \|p - me\|^2$$

le qui assure principalement si p est plus proche de la moyenne de q que de celle de c .

leci nous amène à présenter la procédure algorithmique

PHS 1 : Sélectionner une partition initiale des m individus en k classes
calculer les moyennes et la valeur du critère.

PHS 2 : Choisir le candidat p appartenant à l'une des classes, soit c .

PAS 3 : $\Delta_i m_e = 1$ on va au pas 6
 sinon on calcule

$$e_j = \frac{m_j}{m_j + 1} \parallel p - m_j \parallel \quad m_j \neq e$$

$$= \frac{m_e}{m_e - 1} \parallel p - m_e \parallel \quad m_j = e$$

PAS 4 : Transférer p de la classe le dans la classe e_j si

$$e_j < e_j \text{ quelque soit } j.$$

PAS 5 : Réajuster la valeur du critère, le moyennant de e et q

PAS 6 : Δ_i le critère n'a pas changé après m essais on arrête sinon on reprend au pas 2.

B La méthode des surfaces

Considérons M un ensemble de m objets x_1, x_2, \dots, x_m que l'on désire partitionner en k classes disjointes C_1, C_2, \dots, C_k . Jusqu'à présent, les méthodes proposées visent à regrouper les points les plus similaires, les plus rapprochés. Elles demandent donc toutes le choix d'un indice de similarité, de dissimilarité ou d'une distance. Cette nouvelle méthode ne suppose de regrouper les m points en k classes disjointes telles que la norme des surfaces des enveloppes convexes de chacune des k classes soit minimale. Le principe est donc de trouver k classes telles que :

$$w(a_1, a_2, \dots, a_k) = \sum_{i=1}^k S(a_i)$$

avec $S(a_i)$ la mesure de surface de l'enveloppe convexe de a_i

soit minimal.

Un grand intérêt de cette méthode est qu'elle est justifiée par un modèle probabiliste.

Pour fixer le modèle avec lequel on travaille, il faut faire quelques hypothèses :

- Les variables aléatoires qui comptent les membres de points dans des régions disjointes sont indépendantes.
- Le nombre moyen de points dans chaque région est proportionnel à la mesure de l'espace de cette région.

On évite ainsi un seul processus ponctuel qui

adapte à ces conditions : c'est le processus de Poisson rationnaire. Le modèle, qui apparaît bien sûr

comme restrictif n'a pas à faire comme les hypothèses, est néanmoins d'intérêt assez général.

On effectue :

- Il est la limite (pour des conditions assez générales) des nombres de processus ponctuels indépendants

- Conditionnellement au fait que n points aléatoires engendrés par le processus se trouvent dans une même région, ces points y sont distribués

indépendamment et uniformément. Ceci permet d'exprimer les répartitions des plus aléatoires.

Le nom de ce processus vient de ce que, pour chaque région A , la variable aléatoire de comptage pondée la distribution poissonnienne dont le paramètre est la mesure de Lebesgue de cette région.

Signifions encore que ce modèle peut n'être pas un mélange de densités. Ici veut dire que trouver k densités revient à trouver les k densités présentes dans l'échantillon.

En effet, si un point x appartient à H , il sera dans G avec une probabilité p_j et il y sera distribué suivant la densité $f(x|g)$ avec $p_j = \frac{m(g)}{m(M)}$

et $f(x|g) = \frac{1}{m(g)} \pi_g(x)$

où $m(M)$ est la mesure de M , qui est la somme des mesures (π_i) $i=1, \dots, k$ et $\pi_g(x)$ la fonction indicatrice de G au point x .

Autant ce modèle, la vraisemblance devient $f_M(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p_j f(x_i|g)$

Le domaine M , pour lequel la vraisemblance est maximale, est fourni ceux qui contiennent tous les points, ceux dont la mesure de Rényi est minimum. Si on n'impose pas de contrainte supplémentaire sur M , on constate qu'il existe beaucoup de solutions possibles à ce problème. Cependant l'hypothèse la plus facile qui rende le domaine estimable est la suivante des classes C_i (Rosen [2]).

à une position du nuage de points en R nous associons cependant des enveloppes convexes disjointes correspond

toute une famille d'estimateurs : il suffit de trouver R convexes disjoints contenant chacun un des nous-ensembles. Cependant, pour chaque position, la vraisemblance possède un maximum local continue par les enveloppes convexes des nous-ensembles.

$$= \frac{1}{m(M)^m} \prod_{i=1}^m \prod_{j=1}^m m(g_j(x_i))$$

$$= \prod_{i=1}^m \left(\sum_{j=1}^m \frac{1}{m(M)} \prod_{g_j(x_i)} \right)$$

$$= \prod_{i=1}^m \left(\sum_{j=1}^m \frac{m(g_j)}{m(M)} \prod_{g_j(x_i)} \right)$$

le maximum global de la vraisemblance est donc atteint par la partition pour laquelle la somme des mesures de Rényi des enveloppes convexes de R sous-ensembles est minimale. Pratiquement, si l'espace de base est \mathbb{R} , on cherche les k intervalles disjoints contenant tous les points et tels que la somme de leurs longueurs soit minimale. Dans \mathbb{R}^2 (ou \mathbb{R}^3), on essaie de trouver les k groupes de points tels que la somme des aires (volumes) de leurs enveloppes convexes disjointes soit minimale.

En ce qui concerne l'algorithme, il est une généralisation multidimensionnelle de l'algorithme de Fisher.

aj conditions d'application

- les échantillons sont des convexes
- chaque échantillon c_i , se voit attribuer une mesure $S(c_i)$ (dans \mathbb{R} , la longueur de l'intervalle, dans \mathbb{R}^2 , la surface, dans \mathbb{R}^3 , le volume...)
- et l'erreur de la partition $P(m, k)$ de m objets en k classes. est la somme de ces mesures

$$W(a, c_1, \dots, c_k) = \sum_{i=1}^k S(c_i)$$

b) l'ensemble même de l'algorithme est le lien entre les partitions optimales en k et $(k-1)$ classes. Si on dispose de la meilleure partition en k classes et qu'on supprime la dernière, alors les $(k-1)$ classes restantes doivent former la meilleure partition des points restants en $(k-1)$ classes.

c) la procédure de programmation dynamique est la suivante :

PA5.1 : Calculer les valeurs du coût U_i pour les partitions optimales de i points en

deux classes ($2 \leq i \leq m$)

Pratiquement il suffit de tracer tous les droites passant par deux points du nuage et de calculer la somme des axes des enveloppes convexes des deux classes.

(Cela donne que deux polygones convexes sont toujours séparés par une droite tangente aux deux).

PAS 2 : Pour tout l ($3 \leq l \leq k$) calculer les valeurs du critère pour les positions optimales des l points en l classes

Prenons $l=3$; il suffit de fixer un point de l'enveloppe convexe globale. Le point appartenant forcément à l'une des classes, soit A. On peut ensuite facilement identifier tous les convexes contenant ce point. En effet il existe toujours une tangente commune à la classe A et au deuxième convexe, soit B. Cette droite passe par un point de A et un point de B. Il en va de même pour A et C. On peut alors chercher, pour chaque convexe inclé, la meilleure position des points vérifiant en deux convexes diagonaux entre eux et diagonaux du premier

PAS 3 la position optimale est trouvée à partir de la table des valeurs du critère w en prenant d'abord pour classe C_k , celle qui minimise $w(a, a_1, \dots, a_k) = S(C_k) + w(a, C_1, \dots, C_{k-1})$.

CONDITIONS D'ADMISSIBILITE POINT DE VU THEORIQUE

Choisir une bonne procédure de classification est devenu un véritable problème, vu la myriade de méthodes proposées. Fisher et Van der Veer de déterminer une série de conditions que devrait vérifier une procédure de classification pour être appelée "admissible". Leur approche, motivée par la théorie de la

décision est une des seules qui compare directement les méthodes et de façon qualitative. Nous allons d'abord présenter les neuf conditions, ensuite vérifier leur validité pour les procédures suivantes : méthodes hiérarchiques

du lien minimal
du lien maximal
des centres de gravité
de Ward
méthodes non hiérarchiques

optimisant un critère : l'erreur de la

forme des courbes

de réallocation : Hill (voir leur ouvrage

(Kremer)

et les adapter à la méthode des surfaces.

§1 Présentation des conditions d'admissibilité

Condition 1

Supposons que les n individus x_1, x_2, \dots, x_n aient été ordonnés de telle façon que les classes produites n'échangent pas la forme

$$G = \{x_1, x_2, \dots, x_{m_1}\}, G' = \{x_{m_1+1}, x_{m_1+2}, \dots, x_{m_1+m_2}\}, \dots$$

$$G^k = \{x_{m_1+\dots+m_{k-1}+1}, \dots, x_{m_1}\}.$$

Il faut $\{y_1, y_2, \dots, y_m\}$ une permutation quelconque de $\{x_1, x_2, \dots, x_n\}$. Définissons les ensembles suivants :

$$G' = \{y_1, y_2, \dots, y_{m_1}\}, G'' = \{y_{m_1+1}, y_{m_1+2}, \dots, y_{m_1+m_2}\}, \dots$$

$C_k' = \{y_{m_1+1}, \dots, y_{m_1+m_k+1}\}$ et appelons C_1', C_2', \dots
 C_k' une image de C_1, C_2, \dots, C_k .

Une classification C_1, C_2, \dots, C_k est dite administrable par rapport aux images n' d n' existe pas d'image qui lui soit uniformément meilleure dans le sens

suivant : (a) $d(x_i, y_j) \geq d(y_i, y_j)$ lorsque le i ième et le j ième points sont dans la même classe.

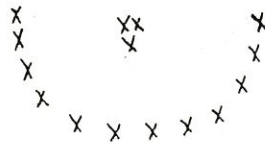
(b) $d(x_i, y_j) \leq d(y_i, y_j)$ lorsque le i ième et le j ième points sont dans des classes différentes avec une inégalité stricte pour au moins une paire d'indices

Une seconde forme d'administrabilité concerne la forme géométrique des classes.

Condition 2

Une procédure est dite convexe-administrable si les enveloppes convexes des k classes produites sont deux à deux disjointes

cette propriété est quelquefois trop exigeante en ce
 sens qu'elle élimine des classifications raisonnables
 telles que le montre les points ci-dessous



Les classes raisonnables sont à constituer des
 points de l'axe de ceale et à constituer
 des points ne situant au centre.
 Pourtant, ces deux classes ne vérifient pas
 la propriété de convergence administrative
 Il est donc nécessaire d'introduire une
 condition plus forte

condition 3 :

Définissons d'abord l'arbre de longueur minimale
 (ou linéaire) L_A d'un ensemble A .
 On appelle arbre L_A un graphe connexe sans cycle
 dans A .

Un graphe est dit connexe si toute paire de points distincts de H , est reliée par une chaîne de H .

Une chaîne reliée x_1, \dots, x_m dans H est une suite d'arêtes de H , $x_1 x_2, x_2 x_3, \dots, x_{m-1} x_m$, telle que

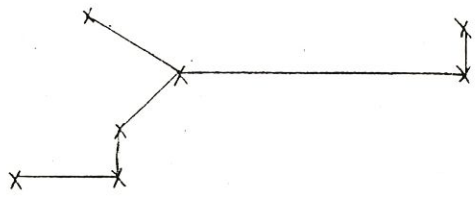
chaque arête $x_i x_{i+1}$ ($1 \leq i < m$) ait un sommet en commun avec l'arête qui la précède et

l'arête nommée avec l'arête qui la suit.

Un arête de longueur minimale est un arête qui minimise la somme des longueurs des arêtes, c'est

$$\sum d(x_i, y) \quad x_i, y \in H$$

Prenons un exemple



Une procédure est dite connexe - admissible si les autres de longueur minimale des le classe produits sont disjoints.

Remarquons qu'une procédure connexe admissible est connexe - admissible.

L'administrabilité par rapport aux proportions enaie
 d'exprimer l'idée que dans certaines applications
 l'aspect géométrique est plus important que la
 densité des points se trouvant dans les classes.

Condition 4.

Une procédure est dite administrable par rapport aux
 proportions de points si une duplication d'un ou
 plusieurs points ne modifie pas les frontières des
 classes.

Condition 5.

Une procédure est dite administrable par rapport aux
 proportions des classes si une duplication de chaque
 classe un nombre arbitraire de fois ne modifie pas
 les frontières des classes.
 Nous eutons dans par doubles une classe, doubles
 tous les éléments d'une même classe un même
 nombre de fois.

Condition 6

Si une procédure donne k classes et si nous enlevons tous les points d'une classe, soit g , alors la procédure est dite admissible par rapport à l'omission de classe si lorsque elle est appliquée aux points restants, afin d'obtenir $(k-1)$ classes, elle donne les mêmes classes à l'exception de g .

Ils devons également que la classification soit indépendante de l'échelle de mesure. Il est donc intéressant de savoir si une transformation monotone de la matrice des distances a une influence sur la classification des données.

Condition 7

Une procédure est dite admissible par rapport aux transformations monotones si une transformation monotone appliquée à chaque élément de la matrice des distances ne change pas la classification. Par transformation monotone, nous entendons toute fonction strictement croissante telle que $f(0) = 0$.

Quand des classes bien séparées sont venues à nous, nous allons d'abord définir ce que sont des données bien structurées en R groupes et en arbre.

1. Des données sont dites bien structurées en R groupes s'il existe une classification C, A, \dots, C telle que toutes les distances intra-classes soient inférieures à toutes les distances inter-classes.

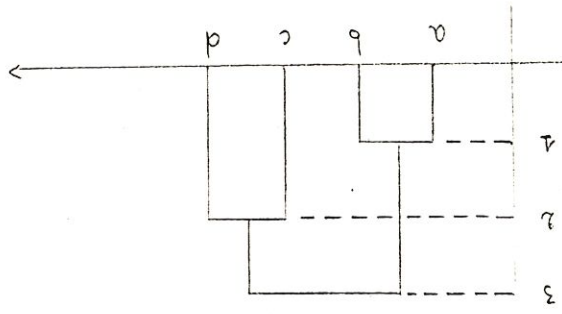
Nous entendons par distance intra-classes, la distance entre deux points d'une même classe. et par distance inter-classes, la distance entre deux points appartenant chacun à des classes différentes. Les données sont donc bien structurées en R groupes si le plus grand des diamètres des classes est inférieur à la plus petite distance inter-classes. (le diamètre étant la distance entre les points les plus éloignés de la classe)

Cette définition de données bien structurées en R classes est donc indépendante de la procédure

2. Des données sont dites bien structurées en arbre si elles ont une structure d'arbre exact, c'est à dire, si on peut reconstituer la matrice des distances des données originales par la seule connaissance de l'arbre hiérarchique.

Examinons de plus près cette définition :

à partir d'une hiérarchie (H, f) on peut construire beaucoup de matrices de distances. Prenons un exemple :



Au plus deux valeurs exactes sont connues dans la matrice des distances, il n'agit de $d(a, b) = 1$ et $d(c, d) = 2$. On peut toujours, connaître à partir de la hiérarchie, un indice de divisibilité σ tel que $\sigma(x_k, x_l) = \min_y \{ f(y) \mid x_k, x_l \in R_y \}$.

Dans notre exemple, cela signifie que

$$\sigma(a, b) = 1$$

$$\sigma(a, c) = \sigma(a, d) = \sigma(b, c) = \sigma(b, d) = 3$$

$$\sigma(c, d) = 2$$

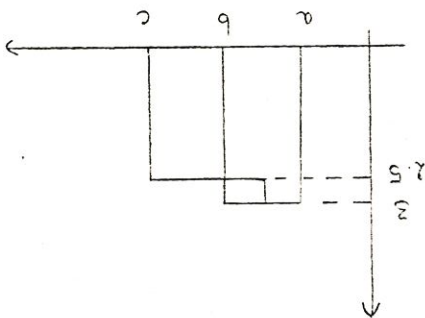
On est donc déjà amené à émettre une hypothèse pour avoir des données bien structurées en arbre :

Hypothèse 1 : $\forall x, y \in M : d(x, y) = \sigma(x, y)$

$$= \min_h \{ f(h) \mid x, y \in h \}$$

Il reste encore un problème, pour cela considérons la

hiérarchie suivante



f indice de dissimilitude, mesuré de la forme

$$\sigma(a, b) = \sigma(a, c) = \sigma(b, c) = 2.5$$

Hypothèse 2 : $\forall h_1, h_2 \in H$ si $h_1 \subset h_2$

$$\text{alors } f(h_1) < f(h_2)$$

ie H est une hiérarchie induite

On est maintenant amené à définir correctement ce que signifie des données bien structurées en arbre (formulation de J.-H. Haultegam [1])

Une matrice de dimensionnalité ① a une structure d'arbre exact H , ou M si pour une fonction f (H, f) est une hiérarchie induite et $d(x_i, x_j) = \min_{R \in H} \{ f(R) \mid x_i, x_j \in R \}$

Remarquons que dans ce cas

σ est une distance ultramétrique.

en effet :

Rappelons que σ est défini à partir de l'hiérarchie induite $\sigma(x_i, x_j) = \min_{R \in H} \{ f(R) \mid x_i, x_j \in R \}$. σ est une ultramétrique n'implique les 3 axiomes

$$\sigma(x_i, x_j) = 0 \Leftrightarrow x_i = x_j$$

en effet $x_i = x_j \Rightarrow f(R) = 0$ car R ne contient

qu'un seul élément

$$\Rightarrow \sigma(x_i, x_j) = 0$$

$\sigma(x, y) = 0 \Rightarrow \exists h = x, y \text{ tq } f(h) = 0$
 $\Rightarrow h$ ne contient qu'un seul élément
 $\Rightarrow x = y$
 $\sigma(x, y) = \sigma(y, x)$
 . l'inégalité ultramétrique
 $\forall x, y, z : \sigma(x, y) \leq \max \{ \sigma(x, z), \sigma(z, y) \}$
 considérons trois points quelconques x, y, z et
 supposons que $h(x, y)$ soit le plus des trois
 valeurs : $f(h(x, y)) \leq f(h(x, z))$
 $\leq f(h(z, y))$
 supposons aussi que $f(h(x, z)) \leq f(h(x, y))$
 alors nécessairement appartenir à $h(x, z)$
 car $h(x, y) \cap h(y, z) \neq \emptyset$
 $\Rightarrow h(x, y) \subset h(y, z)$
 ou $h(y, z) \subset h(x, y)$
 or $f(h(x, y)) \leq f(h(y, z))$
 cela implique que $h(x, y) \subset h(y, z)$
 comme le point $h(y, z)$ contient y, z et x
 et qu'il est strictement plus bas que $h(x, y)$
 on a nécessairement $h(x, z) \subset h(y, z)$ à
 la même hauteur.

nous aurons comment trouver une ultramétrique à partir d'une hiérarchie induite. Sachant à quelle condition une distance est une ultramétrique, nous allons nous intéresser à la construction de la hiérarchie lui correspondant

en effet : tous les triangles isocèles, avec la base plus petite que les côtés vérifient l'inégalité ultramétrique. Considérons un triangle quelconque (x_i, y_j, x_k) et prenons le plus grand côté. Par l'inégalité ultramétrique il est plus petit que les deux autres, mais comme c'est le plus grand, il est donc égal à l'un des deux autres.

Une condition nécessaire et suffisante pour qu'une distance soit ultramétrique sur M est que tous les triangles de M soient isocèles, avec la base plus petite que les côtés

cela implique que $\sigma(x_i, x_k) = \sigma(x_j, x_k)$ de plus $\sigma(x_i, y_j) \leq \sigma(x_i, x_k)$ les trois points forment donc un triangle isocèle avec la base plus petite que les côtés. l'inégalité ultramétrique est donc bien vérifiée. On peut faire le même raisonnement pour tous les triangles de M , on en déduit que σ est une ultramétrique

Construction d'une hiérarchie indexée à partir d'une ultramétrique.

Soit δ une ultramétrique

(H, δ) est la hiérarchie indexée associée à δ avec

H l'ensemble des objets d'équivalence de relation

$R_\alpha, \alpha \in \mathbb{R}^+$ telles que

$$\alpha R_\alpha x' \Leftrightarrow \delta(x, x') \leq \alpha \quad \forall x, x' \in M.$$

$$H = \{ P_\alpha, \alpha \in \mathbb{R}^+ \} \text{ avec } P_\alpha = \{ x, x' \in M \mid \delta(x, x') \leq \alpha \}$$

l'application $f: H \rightarrow \mathbb{R}^+$ est définie par :

$$f(h) = \inf \{ \alpha \mid \alpha \in \mathbb{R}^+, h \in P_\alpha \}$$

Pour que cette construction ait un sens, il faut montrer que R_α est une relation d'équivalence et que (H, f) est une hiérarchie indexée.

* R_α est une relation d'équivalence.

c'est une relation

$$\begin{aligned} \text{réflexive} \quad & \alpha R_\alpha x \Leftrightarrow \delta(x, x) \leq \alpha \quad \text{or } \delta(x, x) = 0 \\ \text{symétrique} \quad & \alpha R_\alpha x' \Leftrightarrow \delta(x, x') \leq \alpha \Leftrightarrow \delta(x', x) \leq \alpha \\ & \Leftrightarrow x' R_\alpha x \end{aligned}$$

$$\text{transitive} \quad \alpha R_\alpha x' \text{ et } x' R_\alpha x'' \Rightarrow \delta(x, x'') \leq \alpha$$

$$\Leftrightarrow \max(\delta(x, x'), \delta(x', x'')) \leq \alpha \Rightarrow x R_\alpha x''$$

induite

Il reste à montrer que (H, f) est une hiérarchie

d'ou $x' \in h'$ et donc $h \subset h'$

ce qui implique $\delta(x, x') \leq \alpha < \alpha'$

on a mécaniquement pour tout $x' \in h$, $x R_\alpha x'$

Supposons que $\alpha < \alpha'$ et prenons $x \in h \cap h'$

h (respectivement h')

tel que l'une des classes d'équivalence induit soit

soit α (respectivement α') le plus petit réel positif

3) $\forall h, h' \in H$ on a : $h \cap h' \neq \emptyset \Rightarrow h \subset h'$ ou $h' \subset h$

qui est connue $\alpha' R_\alpha$

la classe contenant tous les éléments de M

en effet pour $\alpha = \max \delta(x, y)$ il m'y a que

$M \in H$

on $\delta(x, y) \leq 0 \Leftrightarrow x = y$

Classes induites par R_α est formé des singletons

en effet pour $\alpha = 0$, la partition R_0 des classes d'équiva.

" $\forall y \in M$ $\{y\} \in H$

est bien une hiérarchie. Pour cela il faut vérifier que

d'équivalence définies par les relations d'équivalence R_α ,

maintenant que H , l'ensemble des classes

* (H, f) est une fonctionnelle indé-

montrons d'abord que $f(h) = \min \{ \alpha \mid \alpha \in \mathbb{R}^+, h \in P_\alpha \}$
autrement dit que $h \in P_{f(h)}$

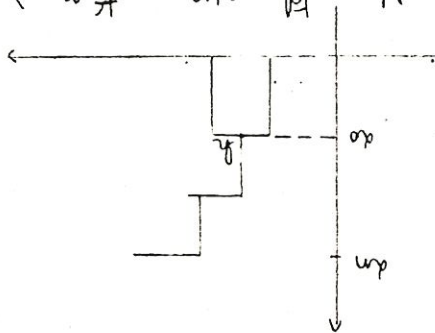
. si h est réduit à un seul élément.

on a $f(h) = 0$ et $h \in P_0$ donc $h \in P_{f(h)}$

. si h contient plus d'un élément.

soit $\alpha_0 = \inf \{ \alpha \mid \alpha \in \mathbb{R}^+, h \in P_\alpha \}$ et considérons

une suite décroissante $\alpha_n \rightarrow \alpha_0$.



il existe N tel que $\forall m > N$ $h \in P_{\alpha_m}$

et donc par définition de P_{α_m}

$\forall (x, x') \in h$ et $\alpha'' \notin h$ on a $\delta(x, x') \leq \alpha_n$
et $\delta(x, \alpha'') > \alpha_n$

en passant à la limite on obtient

$\delta(x, x') \leq \alpha_0$ et $\delta(x, \alpha'') > \alpha_0$

donc tous les éléments de h sont dans une classe
d'équivalence de P_{α_0} et cette classe ne contient

aucun élément qui n'est pas dans h .

d'où $h \in P_{\alpha_0} = P_{f(h)}$.

Récapitulons : nous avons défini des données bien structurées en arbre, et nous avons vu qu'à partir d'une hiérarchie on peut construire une ultramétrique et qu'à partir d'une ultramétrique on peut retrouver une hiérarchie induite.

capd

Il faut $x, h, h' \in H$. $h \subset h' \Rightarrow f(h) < f(h')$
 en effet : $f(h) = \min \{ \alpha \mid \alpha \in \mathbb{R}^+, h \in P_\alpha \}$
 $f(h') = \min \{ \alpha \mid \alpha \in \mathbb{R}^+, h' \in P_\alpha \}$
 soit $x \in h' \setminus h$, on a nécessairement
 $\delta(x, x') > f(h)$ $\forall x' \in h$ (sinon $x \in h$)
 et $\delta(x, x'') < f(h')$ $\forall x'' \in h'$ ($h \subset h'$)
 on a donc $f(h) < \delta(x, x') < f(h')$
 d'où $f(h) < f(h')$.

Il faut $f(h) = 0 \Leftrightarrow h$ ne contient qu'un seul élément
 en effet : $f(h) = 0 \Leftrightarrow \min \{ \alpha \mid \alpha \in \mathbb{R}^+, h \in P_\alpha \} = 0$
 $\Leftrightarrow h \in P_0$
 $\Leftrightarrow \forall x, x' \in h, \delta(x, x') = 0$
 $\Leftrightarrow h$ ne contient qu'un seul élément

l'application f doit satisfaire aux deux propriétés suivantes

On peut même démontrer qu'il existe une bijection entre les ultramétriques et les réseaux induits [annexe 2].

Nous allons maintenant définir les deux dernières conditions d'admissibilité:

Condition 8

Une procédure est bien structurée en k classes si lorsqu'elle est appliquée à des données bien structurées en k groupes, elle redonne ces k classes.

Condition 9

Une procédure est bien structurée en classe si lorsqu'elle est appliquée à des données bien structurées en classe, elle redonne la même classe hiérarchique.

§ 2 Vérification des conditions

A. Méthodes hiérarchiques

Les procédures du lemme minimal et du lemme maximal sont administrées par rapport aux images.

Considérons une application bijective $f: M \rightarrow M$ qui réordonne les éléments de M et telle que :

si x, y sont dans la même classe :

$$(a) \quad d(f(x), f(y)) \leq d(x, y)$$

si x et y sont dans des classes différentes :

$$(b) \quad d(f(x), f(y)) \geq d(x, y)$$

montrons que ces inégalités sont des égalités.

Nous allons d'abord démontrer, par récurrence sur les niveaux de la hiérarchie que si au niveau i les éléments sont

$$\{c_1(i), \dots, c_{k(i)}(i)\} \text{ alors } [f(c_1(i)), f(c_2(i)), \dots, f(c_{k(i)}(i))] = [c_1(i), c_2(i), \dots, c_{k(i)}(i)]$$

Du niveau 0

évident car les énoncés sont formés de singuliers.

Du niveau 1

on suppose l'hypothèse de récurrence vérifiée.

Du niveau $(i+1)$

Supposons que si on suppose deux énoncés, soient $c(i)$ et $q(i)$. Cela signifie que

$$\delta(c(i), c(i)) \leq \delta(c(i), m(i))$$

$$1 \leq m, n \leq f(i)$$

alors on suppose également $f(c(i))$ et $f(q(i))$.

$$\delta(c(i), c(i)) \leq \delta(c(i), m(i))$$

\Downarrow

$$\delta(f(c(i)), f(q(i))) \leq \delta(f(c(i)), f(m(i)))$$

si on considère la distance du bien minimal

$$(c(i), q(i)) = \min_{\substack{re \in c(i) \\ qf \in q(i)}} d(re, qf)$$

$$\text{or } d(re, qf) \geq d(f(re), f(qf)) \text{ par (a)}$$

$$\text{et donc } \min_{\substack{re \in c(i) \\ qf \in q(i)}} d(re, qf) \geq \min_{\substack{re \in c(i) \\ qf \in q(i)}} d(f(re), f(qf))$$

on peut conclure que $d(x_i, y_j) = d(f(x_i), f(y_j))$ pour x_i, y_j appartenant à des classes différentes.

$$\sum_{i=1}^k \sum_{j=1}^l d(x_i, y_j) = \sum_{i=1}^k \sum_{j=1}^l d(f(x_i), f(y_j))$$

de manière analogue pour des classes différentes

avec (a) on peut conclure que $d(x_i, y_j) = d(f(x_i), f(y_j))$ pour x_i, y_j appartenant à la même classe.

$$\sum_{i=1}^k \sum_{j=1}^l d(x_i, y_j) = \sum_{i=1}^k \sum_{j=1}^l d(f(x_i), f(y_j))$$

pour des classes différentes

on peut donc dire que :

$$[f(q(c_1)), f(q(c_2)), \dots, f(q(c_{k+1}))] = [q(c_1), q(c_2), \dots, q(c_{k+1})]$$

ceci prouve que

du lien maximal.

on peut procéder de la même manière pour la distance

$$\delta(f(q(c_1)), f(q(c_2))) \leq \delta(q(c_1), q(c_2))$$

$$\delta(q(c_1), q(c_2)) \leq \delta(c_1, c_2) \quad 1 \leq m, m \leq k(c)$$

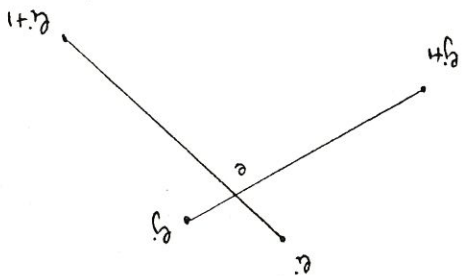
ce qui implique que :

La procédure du lien minimal est convergente - admissible

Nous devons démontrer que les chaînes de longueur minimale des classes sont à 2 diagonales.

Procédons par l'absurde et supposons qu'il existe

L_a et L_g tels que $L_a \cap L_g = \{e\}$ et $i \neq j$



Supposons e_i le point appartenant à L_a

le plus proche de e

e_j le point appartenant à L_g

le plus proche de e

Nous savons par construction de la chaîne L_g et

par la procédure que e_j est le point le plus proche de e_{j+1}

$$d(e_j, e_{j+1}) \leq d(e_j, e_i)$$

$$\text{or } d(e_j, e_{j+1}) = d(e_j, e) + d(e, e_{j+1})$$

Nous pouvons donc déduire que

$$d(e_j, e) + d(e, e_{j+1}) \leq d(e_j, e_i)$$

$$\leq d(e_j, e) + d(e, e_i)$$

$$d(e, e_{j+1}) \leq d(e, e_i) \quad (1)$$

longueurs des points suivants

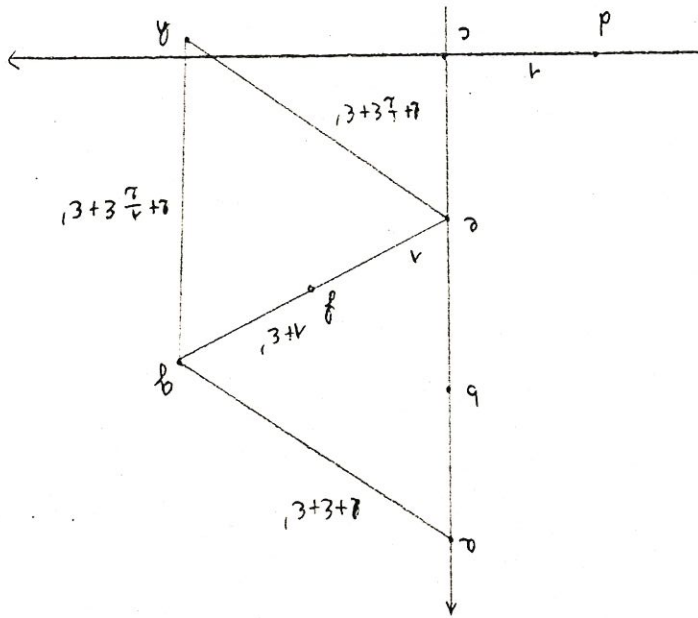
la procédure du lien minimum n'est pas connue.

$d(e_i, y) \leq d(e_i, e) + d(e, y)$
 (non convention)
 (i) non
 $d(e_i, e) \leq d(e_i, e) + d(e, y)$
 (non convention)

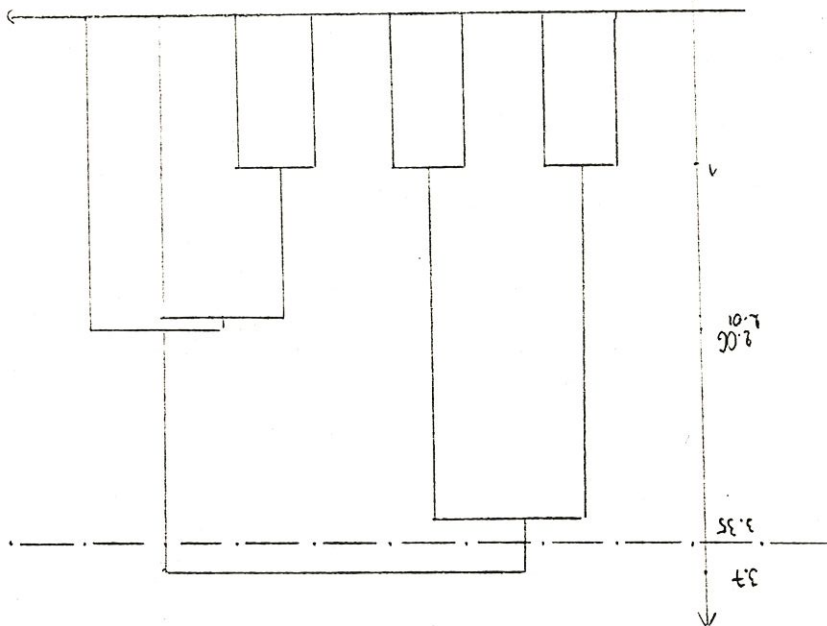
la procédure du lem maximal n'est pas convergente
admissible et donc pas convergente admissible.

considérons les huit points suivants

- a (0, 3.2)
 - b (0, 2.2)
 - c (0, 0)
 - d (-1, 0)
 - e (0, 1.1)
 - f (0.88, 1.57)
 - g (1.77, 2.05)
 - h (1.68, -0.07)
- ($10.0 = 3$, $1.0 = 3$)



la procédure donne la solution suivante.

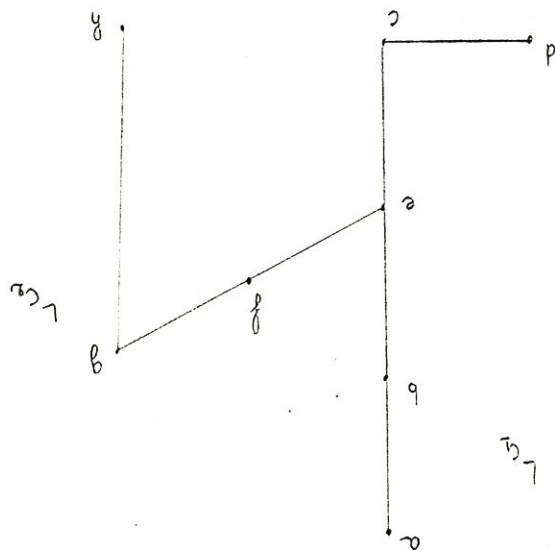


Si on veut une classification en deux classes, on coupe la hiérarchie, et on obtient les deux classes.

$$C_1 = \{a, b, c, d\}$$

$$C_2 = \{e, f, g, h\}$$

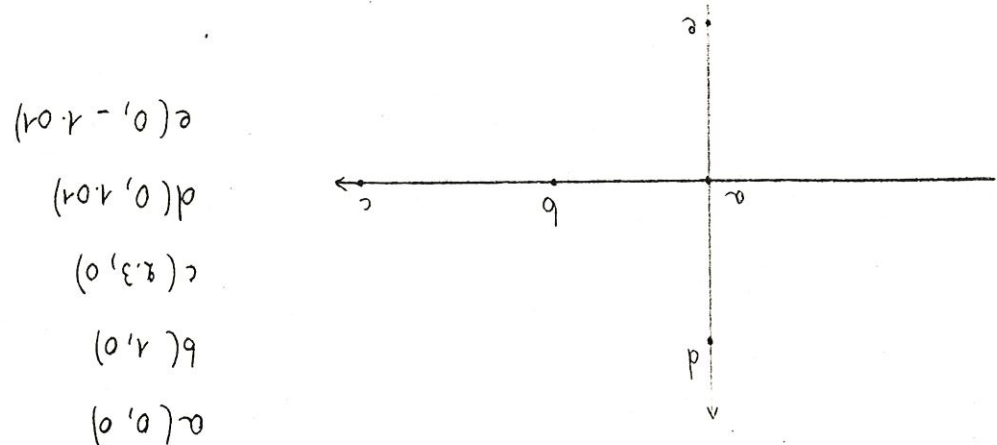
dont les arbres de longueur minimale ne sont pas dirigés



$$L_2 \cup L_1 = \{e\}$$

la réduction des centres de gravité m_i est par conséquent admissible et donc par conséquent admissible.

considérons les points suivants



Preons 1000 fois le point b et 10^5 fois le point c . Appliquons la méthode à ces points. La matrice des distances entre les classes est de la forme suivante

	$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$	$\{e\}$
$\{a\}$	0	1	2.3	1.01	1.01
$\{b\}$	0	0	1.3	1.42	1.42
$\{c\}$			0	2.51	2.51
$\{d\}$				0	2.02
$\{e\}$					0

où i l'élément de la i ème ligne et de la j ème colonne

est la distance entre la $i^{\text{ème}}$ et la $j^{\text{ème}}$ classe.

La distance étant symétrique, on évalue qu'une

partie de la matrice.

au premier niveau, on fusionne les classes les plus rapprochées $\{a\}$ et $\{b\}$.

On réajuste ensuite la matrice des distances, sachant que

$$d(\{a,b\}, c) = d(c, \{a,b\}) = d(c, \{a\}) = d(c, \{b\})$$

	$\{a, b\}$	$\{c\}$	$\{d\}$	$\{e\}$
$\{a, b\}$	0	1.30	1.42	1.42
$\{c\}$		0	1.42	1.42
$\{d\}$			0	1.42
$\{e\}$				0

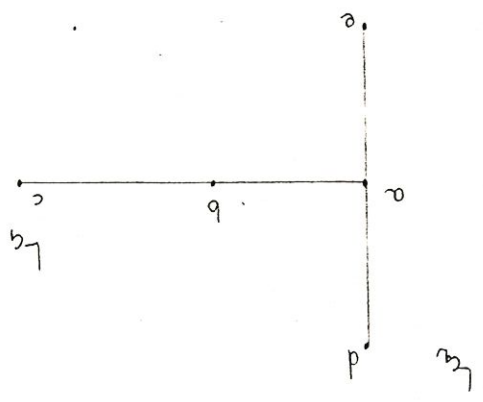
les deux classes les plus proches sont $\{c\}$ et $\{a, b\}$.

Au troisième niveau, la matrice des distances devient

	$\{a, b, c\}$	$\{d\}$	$\{e\}$
$\{a, b, c\}$	0	2.51	2.51
$\{d\}$		0	2.02
$\{e\}$			0

on fusionne $\{e\}$ et $\{d\}$

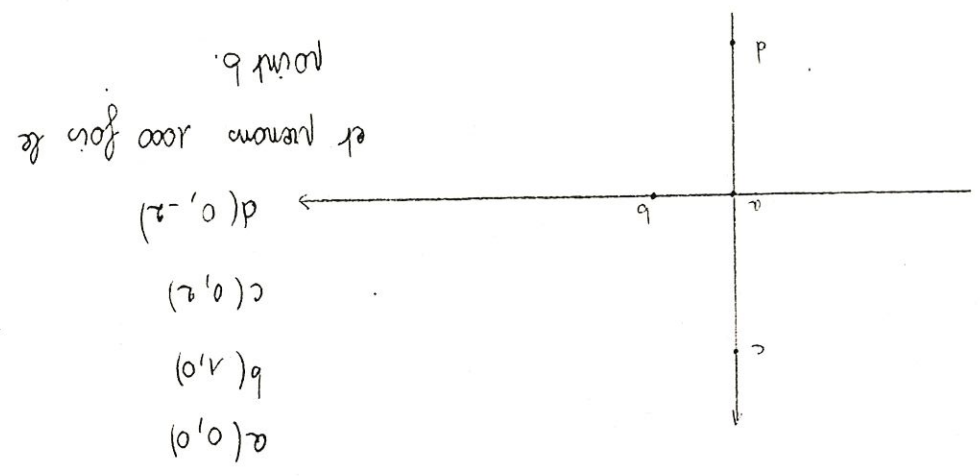
la procédure donne deux classes $a = \{a, b, c\}$ et $c_1 = \{e, d\}$ dont les autres ne sont pas disjointes



$$L_1 \cap L_2 = \{a\}$$

la procédure de Ward n'est pas comme - admissible, et donc pas comme - admissible.

considérons les points suivants



	$\{a, b\}$	$\{c\}$	$\{d\}$
$\{a, b\}$	0	1.23	2.23
$\{c\}$		0	2
$\{d\}$			0

$\{c\}$ et $\{d\}$

Au deuxième niveau, la méthode fusionne les classes

$\{b\}$

Les deux classes les plus rapprochées sont $\{a\}$ et

$$\text{avec } d(a, g) = \frac{m_i m_j}{m_i + m_j} d(\alpha(a), \alpha(g))$$

	$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$
$\{a\}$	0	0.99	1	1
$\{b\}$		0	1.23	2.23
$\{c\}$			0	2
$\{d\}$				0

Au premier niveau, la matrice des distances vaut

La classification finale est donc $C_1 = \{a, b\}$ et $C_2 = \{c, d\}$.
les arbres de longueur minimale du cas deux classes ne
sont pas disjoints.

Remarquons qu'au lieu de doubles nœuds 1000 par un point
on peut considérer 1000 points liés par des nœuds du point
considéré.

Conditions 4-5

administrative par rapport aux proportions des points et
des classes

les procédures du lien minimal et du lien maximal
sont admises par rapport aux proportions des points
et par rapport aux proportions des classes

Comme la distance entre deux classes est la plus
petite (respectivement la plus grande) distance entre
deux points appartenant chacun à une classe différente,
la présence d'un ou plusieurs points doubles n'a pas
aucune influence sur la matrice des distances et donc
aucune sur la classification.

La procédure des centres de gravité n'est pas admissible par rapport aux proportions des points

On peut s'attendre à ce que la procédure ne soit pas admissible par rapport aux proportions des points car la présence d'un point double aura plus de "poids" dans la détermination du centre de la classe et donc la mesure des distances peut être différente et donner des classes différentes.

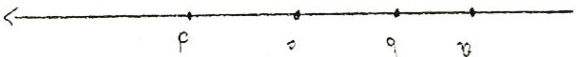
Preons un exemple en considérant les quatre points suivants

$$a = 0$$

$$b = 0.50$$

$$c = 1.16$$

$$d = 1.86$$



La position finale de ces quatre points en deux classes est:

$$C_1 = \{a, b\}, C_2 = \{c, d\}.$$

Si on considère la fois le point b , la classification est

$$C_1^* = \{a, b, c\}, C_2^* = \{d\}$$

telle est bien différente de $[C_1, C_2]$.

la procédure des centres de gravité est admissible

par rapport aux positions des élanes.

Puisque ni on double les points d'une élane, cela ne change pas le centre de gravité, on se doute que la procédure soit admissible par rapport aux positions des élanes.

Uéuon par réuence aux les niveaue de formation des élanes que ni on applique la procédure aux éléments de M ou aux éléments de N plus leur de la élane doublée, soit y , on obtient les mêmes positions en le élanes.

Au niveau 0

évident ;

Au niveau 1

Appelons P_i

l'ensemble des élanes formés à partir de M et avant la création de la élane i

leur implique que les centres de points n'appartenant pas à G . points appartenant à G , soit de C_0 et C sont formés soit de

$$\text{or } \begin{matrix} \text{mun} \\ C_0 \in P_{C_0} \\ C \in P_{C_0} \end{matrix} \quad d(C_0, C) = \text{mun} \quad \begin{matrix} C_0 \in P_{C_0} \\ C \in P_{C_0} \end{matrix} \quad d(C_0, C)$$

$$\begin{matrix} \text{mun} \\ C_0 \in P_{C_0} \\ C \in P_{C_0} \end{matrix} \quad d(C_0, C) = d(C_0, C_0) = d(C_0, C_0) = d(C_0, C_0)$$

$$C_0, C_0 \in P_{C_0}$$

appartenant à P_{C_0} . Autrement dit

on finit au moins deux plans, soient C_0 et C_0

$$\text{montrons que } P_{C_0+1} = P_{C_0}$$

au niveau (C_0+1)

$$P_{C_0} = P_{C_0}$$

l'hypothèse de récurrence devient

(i) plus petit que le niveau de formation de P_{C_0+1} (plane).

création de C_0

partir de $(M \cup G)$ et avant la

l'ensemble des plans formés à P_{C_0}

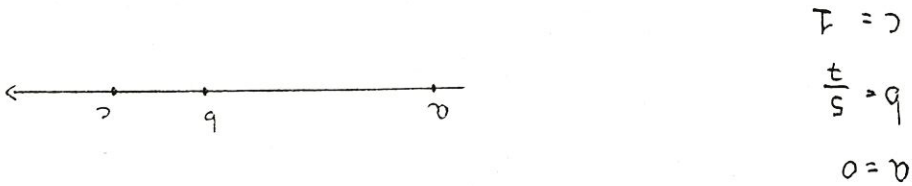
garantir restent les mêmes si on considère P_{i+1} et P_i ou dans P_{i+1} .

Les deux classes fusionnées à l'étape $(i+1)$ de la procédure appliquée à $(H \cup G)$ sont les mêmes que celles fusionnées au cours de la procédure appliquée au M.

La procédure est donc admissible par rapport aux proportions des classes.

La procédure de "Wald" n'est pas admissible par rapport aux proportions de classes.

Considérons les trois points de la droite réelle



que l'on désigne comme en bleu pour la méthode de Wald. Rappelons que

$$\delta(a, g) = \frac{m_x m_y}{m_x + m_y} d(g(a), g(g))$$

La partition en deux classes est fournie de

$$C_1 = \{a\} \text{ et } C_2 = \{b, c\}$$

Si on considère 50 fois les points b et c , la procédure

donne comme classification

$$C_1 = \{a, b\} \text{ et } C_2 = \{c\}$$

(la distance entre $\{b\}$ et $\{c\}$ qui était de $\frac{1}{2} \cdot \frac{2}{3} = 0.4$

$$\text{devient } \frac{50}{100} \cdot \frac{2}{3} = 7.14)$$

La procédure favorise l'union de classes contenant

peu d'éléments.

On peut reprendre le même exemple pour montrer que la procédure m' est non administrative par rapport aux négociations des points.

Condition 6

Administrative par rapport à l'union de classe

Les procédures hiérarchiques sont administratives par rapport à l'union des classes.

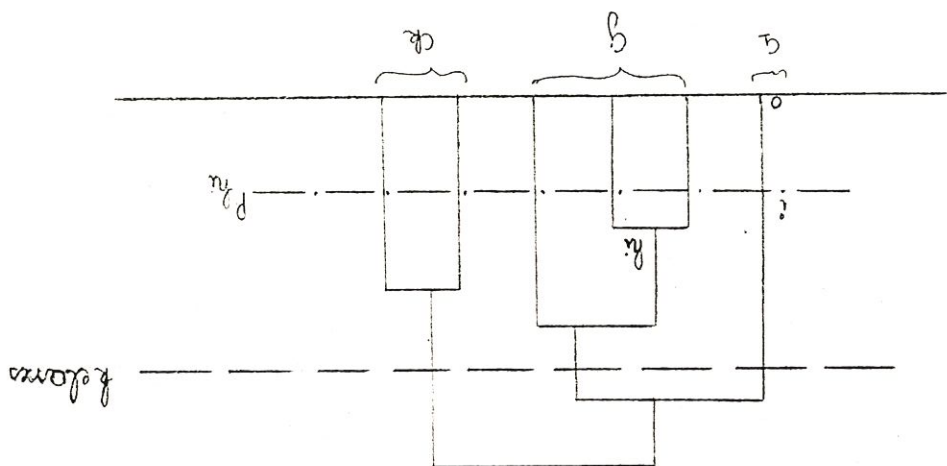
l'union par récurrence sur les niveaux de la

hiérarchie que

si on applique une procédure hiérarchique aux éléments de M afin d'obtenir k classes et que l'on relie tous les éléments d'une classe, soit g ,

ou

si on applique une procédure hiérarchique aux éléments de $(M \setminus g)$ afin d'obtenir $(k-1)$ classes
on obtient la même partition en $(k-1)$ classes.



Appelons P_h la partition formée avant la création de h_i lorsque l'on applique l'algorithme à M .
de h_i lorsque l'on applique l'algorithme
de h_i lorsque l'on applique l'algorithme

Au niveau 0 : évident car la partition est formée de

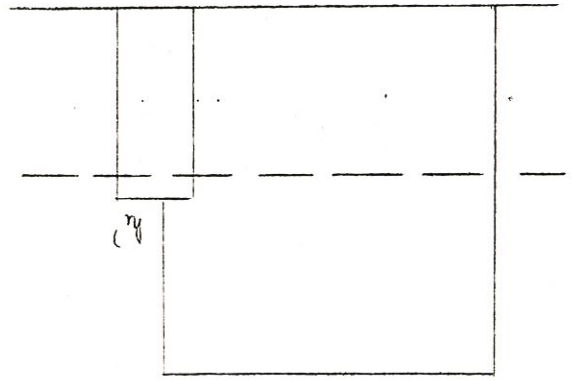
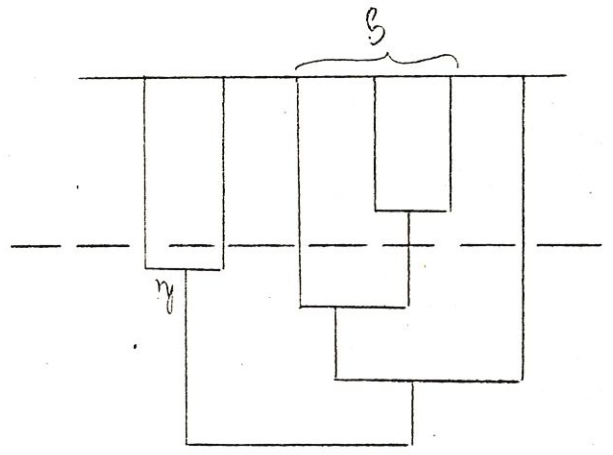
singletons

Au niveau i :

on suppose : $\exists h_i' \mid P_{h_i} \setminus D_i = P'_{h_i'}$

avec $D_i = \{A_i \mid A_i \text{ contient des}$

éléments de $g \text{ et } h_i\}$



Au niveau $(i+1)$

il faut montrer que pour h_{i+1} il existe un palier
tel que $P_{h_{i+1}} \setminus D_{i+1} = P'_{h_{i+1}}$

Différents cas peuvent se présenter :

• si h_i est l'union de deux paliers contenant
des éléments de g alors $h_{i+1} = h_i$

• Si h_i est l'union de deux paliers me

contenant pas des éléments de g

alors h_{i+1} est le palier formé après h_i

• h_i ne peut être la jonction d'un palier conti-
nant des éléments de g et d'un palier m_i en
contenant pas ces nions au niveau de la
jonction des 2 classes, g contiendrait d'autres
éléments

Condition 7

Admissibilité par rapport aux transformations monodromes

Les procédures du part minimum et du part maximum

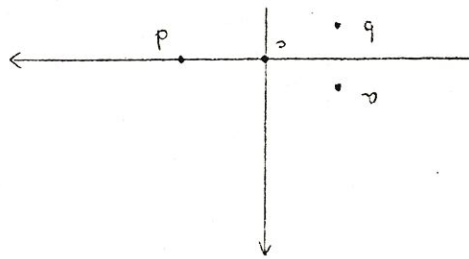
sont admissibles par rapport aux transformations monodromes

Les méthodes ne font pas intervenir les valeurs des distances
mais uniquement leur comparaison. J'ai une transformation
monodrome, les distances les plus petites (les plus grandes) le
représentent. Une transformation monodrome m'a aussi donc comme
influence sur la classification.

La procédure des ventes de gauche n'est pas admissible
 pas rapport aux transformations monochromes.

Donc déterminons les quatre points

$$a \left(-\frac{1}{2} \left\{ (1+\varepsilon)^2 - \frac{1}{4} \right\}^{\frac{1}{2}}, \frac{1}{2} \right) \\
b \left(-\frac{1}{2} \left\{ (1+\varepsilon)^2 - \frac{1}{4} \right\}^{\frac{1}{2}}, -\frac{1}{2} \right) \\
c (0, 0) \\
d \left(1 + \frac{1}{2} \varepsilon, 0 \right)$$



que l'on déduit regroupés en deux classes.
 La matrice des diagrammes est de la forme

	$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$
$\{a\}$	0	1	$1+\varepsilon$	*
$\{b\}$	0	0	$1+\varepsilon$	*
$\{c\}$			0	$1 + \frac{1}{2} \varepsilon$
$\{d\}$				0

$$* = \left[2 + 3\varepsilon + \frac{1}{2} \varepsilon^2 + (1+\varepsilon)^2 - \frac{1}{4} \right]^{\frac{1}{2}}$$

les classes obtenues en appliquant la procédure à ces
 données sont $a = \{a, b, c\}$, $a = \{d\}$

Si on applique une transformation monotone f aux éléments de la matrice des distances, la procédure ne donne pas nécessairement les mêmes classes.

considérons f telle que $f(1) = 1$

$$f(1 + \frac{1}{2}) = 1 + \frac{1}{2}$$

$$f(1 + \frac{1}{3}) = 1 + \frac{1}{3}$$

$$f(1 + \frac{1}{2} + \frac{1}{3}) = 1 + \frac{1}{2} + \frac{1}{3}$$

Les classes obtenues à partir de la nouvelle matrice des distances sont

$$C_1^* = \{a, b\} \text{ et } C_2^* = \{c, d\}$$

différentes de C et C'

La procédure de Ward n'est pas admissible par rapport aux transformations monotones.

Considérons la matrice des distances suivante

	a	b	c	d
a	0	1.14	1.14	2.
b		0	1.	1.1
c			0	1.94
d				0

Il nous devons montrer que lorsqu'on applique ces méthodes à des données bien structurées en arbre, on obtient le même arbre. Si les données sont bien structurées en arbre, cela signifie que δ indice de distance est négative. Or si on peut montrer

des procédures du bien minimal, du bien maximal et de Ward sont bien structurées en arbre

Structure - administrable

Conditions 8-9

différentes de a et c . La procédure n'est donc pas admissible par rapport aux transformations monotones.

$$G^* = \{b, c, d\}$$

la méthode produit les classes $G^* = \{a\}$ et

les autres distances restent invariantes.

$$f(1.4) = 1.8$$

$$f(1.4) = 1.5$$

appliquons la transformation monotone suivante.

$$G = \{a, b, c\} \text{ et } G = \{d\}$$

la procédure donne les deux classes

que ces procédures engendrent des hiérarchies induites, on est certain de résoudre la hiérarchie de départ car il existe une bijection entre les hiérarchies induites et les algorithmiques. Il reste à savoir à quelle condition, on est en présence d'une hiérarchie induite. Rappelons nous la formule de réécriture générale :

$$s(h, h_i \cup h_j) = a_1 s(h, h_i) + a_2 s(h, h_j) + a_3 s(h_i, h_j) + a_4 f(h_i) + a_5 f(h_j) + a_7 |s(h, h_i) - s(h, h_j)|$$

$$\forall h \in P_{h_i \cup h_j}$$

À la hiérarchie H est induite de façon à ce que la hauteur de chaque nœud corresponde à la valeur de l'indice d'aggrégation pour les deux nœuds qui l'ont formé, il existe une condition suffisante sur les coefficients a_1, \dots, a_7 pour assurer l'existence de la hiérarchie induite (H, f) .

$$\delta(h, h_2) \geq \delta(h_1, h_2)$$

$$\delta(h, h_2) \geq \delta(h_1, h_2)$$

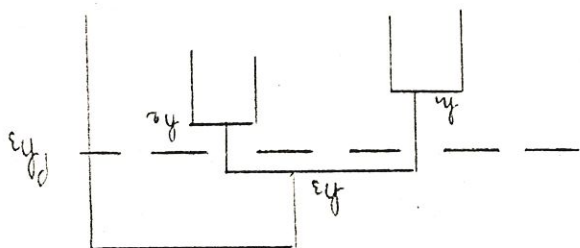
on a maintenant, pour tout $h \in P_{h_3}$ différent

$$\delta(h, h_3) \geq a_1 \delta(h, h_1) + a_2 \delta(h, h_2)$$

puisque a_1, a_2, a_3 sont positifs, on a par la

formule de récurrence générale :

$$a_1 \geq 0$$



à l'étape précédente la formation de h_3

et la partition P_{h_3} obtenue par l'algorithme

considérons un élément $h_3 = h_1 \cup h_2$ de H

Démonstration

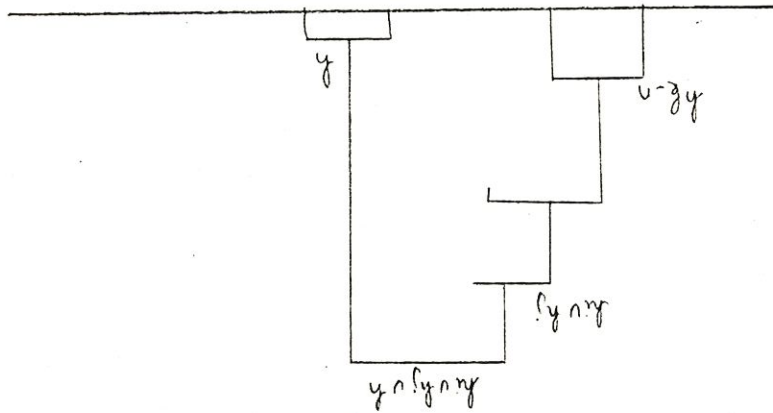
soit

alors (H, f) est une hiérarchie induite ou non

$$a_1 + a_2 + a_3 \geq 1, a_1 \geq 0$$

$$f = 1, 2, 4, 5, 6$$

avec (2) $f(h) \leq f(h_{k-m}) \leq \dots \leq f(h \cup h_j)$



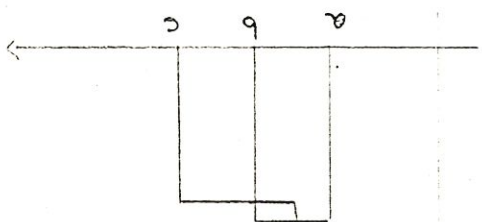
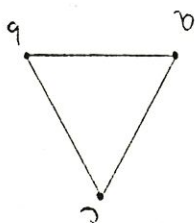
de h . d'ou $f(h_{k-m}) \geq f(h)$

Les implique qu'on a une relation induite.
 Au cours de l'algorithme h est construit avant
 $h \cup h_j$ cela entraîne que $f(h \cup h_j) \geq f(h)$. On effect
 on peut construire une suite de valeurs h_k, \dots, h_{k-n}
 ou h_{k-m} est formé de deux éléments de P_h
 distincts des deux valeurs permettant la formation

on a $\delta(h, h \cup h_j) \geq \delta(h_i, h_j) = f(h \cup h_j)$ (2)

et donc $\forall h \in P_{h_i \cup h_j}$ et $h \neq h_i, h \neq h_j$
 $\delta(h, h \cup h_i) \geq \delta(h, h_j) = f(h \cup h_i)$

on a $\delta(h, h \cup h_i) \geq (a_1 + a_2 + a_3) \delta(h, h_i)$
 si $a_1 + a_2 + a_3 \geq 0$ on obtient



du 3^e côté.

Remarquons que la procédure des centres de gravité peut donner lieu à des inversions. Il suffit d'appliquer la procédure à un triangle isocèle dont la hauteur est un peu plus petite que la longueur

conditions de ce théorème.

On vérifie facilement que les coefficients de la formule de récurse, de la méthode du bien minimal, du bien maximal et de l'aud vérifient les

On a donc bien une hiérarchie induite

$$\text{d'où } f(h) \geq f(h')$$

$$f(h_{i+1}) \geq \max \{ f(h_i), f(h'_{i+1}) \}$$

$$\text{non (5) } f(h_m) \geq \max \{ f(h_{m-1}), f(h'_{m-1}) \}$$

on il existe une suite finie $h = h_1 < h_2 < \dots < h_m = h'$

Il reste à voir que si $h < h' \Rightarrow f(h) \leq f(h')$

$$\text{d'où } f(h \vee h'), h) \geq \max \{ f(h \vee h'), f(h) \} \quad (3)$$

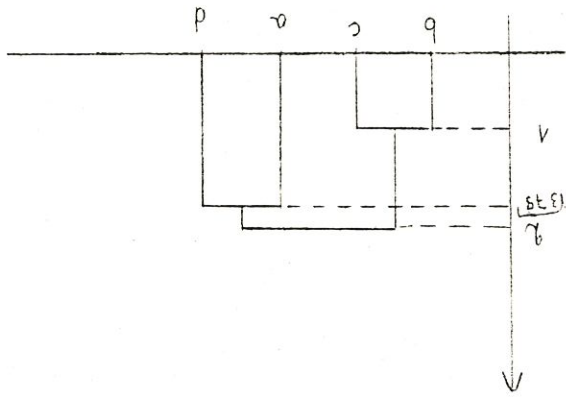
$$\text{on a donc finalement } f(h \vee h'), h) \geq f(h \vee h') \geq f(h)$$

La procédure des centres de gravité m'est pas bien structurée en ordre.

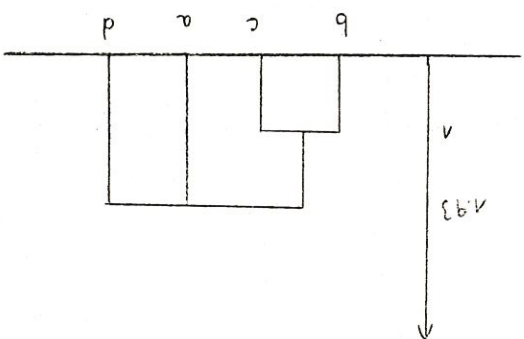
Considérons les quatre sommets d'un tétraèdre dont la matrice des distances est la suivante :

	{a}	{b}	{c}	{d}
{a}	0	2	2	$\sqrt{3.75}$
{b}		0	1	2
{c}			0	2
{d}				0

Les données sont bien structurées en ordre car l'indice de dissimilarité associé à la matrice est une ultramétrique. On peut trouver une fonction f , telle que la hiérarchie soit induite



lorsqu'on applique la procédure des centres de gravité à cet ensemble de points, on ne réobtient pas la même hiérarchie.



La procédure du lien minimal est bien structurée en 2 classes.

hypothèse que les données soient bien structurées en 2 classes b_1, b_2, \dots, b_k .
 $\max_{l=1, \dots, k} \text{diam}(B_l) < \min d(a, b)$ avec $\text{diam}(B_l) = \max d(a_l, y_l)$
 $x_l, y_l \in B_l$

Appliquons la procédure à ces données et regardons si les k classes produites a_1, a_2, \dots, a_k correspondent aux classes b_1, b_2, \dots, b_k . Montrons par récurrence au 1^{er} niveau que les classes ne contiennent pas des éléments de B_i différents.

Au niveau 0

évident car les classes sont des singletons

Au niveau 1

on suppose que $\exists j, \forall a \in a_i, \forall b \in b_j$

Au niveau $(i+1)$

on fusionne au moins deux classes, soit a_i, a_j et montrons que ces classes ne contiennent que des éléments d'un seul B_i .

nous avons que

$$d(a_i, a_j) = d(a_i, a_m)$$

$$\min_{a_m \in a_i} d(a_i, a_j) \leq \min_{a_m \in a_i} d(a_i, a_m) \quad (1)$$

On ne peut avoir $x \in B_c$ et $y \in B_f$

en effet

$$\begin{aligned} \min d(x, y) &> \min d(b_i, b_j) \\ &> \max \text{diam}(B_i) \\ &> \max d(x, y) \\ &> \min d(x, y) \end{aligned}$$

$x \in B_c$
 $y \in B_f$

On ne peut donc fusionner deux classes contenant des points appartenant à des classes bien structurées différentes.

On procède de manière analogue pour démontrer que la procédure du lien maximal est bien structurée en le classe

(il suffit de remplacer (i) par

$$\max d(x, y) \leq \max d(x_m, x_n)$$

$x \in C_c$
 $y \in C_f$
 $x_m \in C_m$
 $x_n \in C_n$

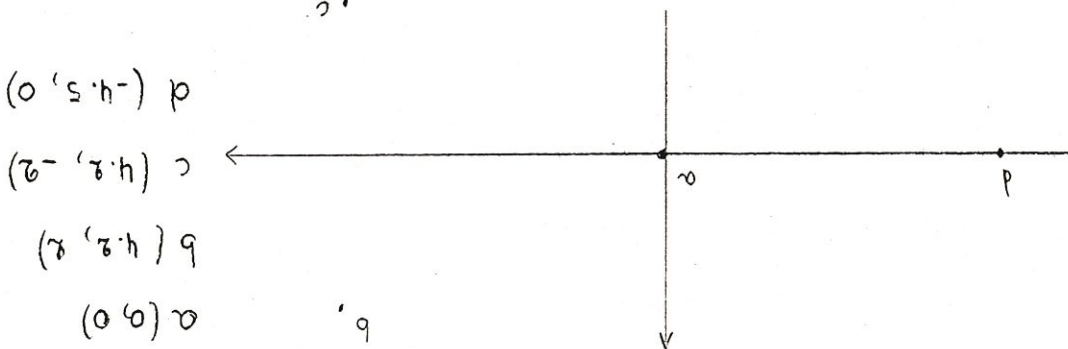
par

$$\max d(x, y) > \min d(b_i, b_j)$$

$x \in B_c$
 $y \in B_f$

La procédure des centres de gravité m'ont pas bien
 effectuée en le classe.

considérons les données suivantes



elles sont. bien effectuées en le classe B_1 et B_2

$$B_1 = \{a, d\}$$

$$B_2 = \{b, c\}$$

en effet le plus grand diamètre (diam (B_1)) = 4.5

est inférieur à la plus petite distance

inter classe (4.65)

Pourtant la procédure ne produit pas les deux

classes. Elle donne $C_1 = \{a, b, c\}$ et $C_2 = \{d\}$

qui ne sont pas bien effectuées en 2.

La procédure de Ward m'ont pas bien effectuée

en le grouper

Reprenons le même exemple que ci dessus mais doublons le point d .
 La procédure donne les 2 classes $a = \{a, b, c\}$ et $a = \{d\}$ différentes du deux classes bien distinctes.

B. Méthodes de l'erreur de la norme du carré et du K. means

Comme la procédure de l'erreur de la norme des carrés optimise globalement le critère

$$f(c_1, c_2, \dots, c_k) = \sum_{j=1}^k \sum_{x_i \in c_j} \|x_i - m_j\|^2$$

et la méthode Hill Climbing leur équivalant (K means).
Cependant, il suffit de montrer qu'une condition est vérifiée pour la méthode locale pour en déduire qu'elle le sera pour la méthode globale.

1. Procédure de l'erreur de la norme des carrés

La procédure est admissible par rapport aux images

considérons une application bijective $f: M \rightarrow M$
et appelons $f(c_i)$ l'ensemble des images des points
de la classe c_i .
 $f(c_i) = \{f(x) \mid x \in c_i\}$

pas convergent

$$\sum_{x \in G} \|x - m_y\|^2 = \sum_{x \in G} \|x - m_x\|^2$$

en effet nous avons que

$$m_y = \text{card}(G)$$

$$\sum_{x \in G} \frac{x}{m_y} = f(m_y)$$

$$\sum_{x \in G} \|x - m_y\|^2 > \sum_{x \in G} \|f(x) - f(m_y)\|^2$$

soit G telle que

les voudrait dire qu'il existe au moins une classe,

avec au moins une inégalité stricte

dans des classes distinctes.

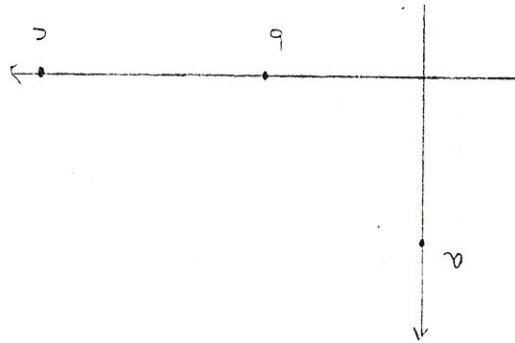
$$(2) \quad d(x, y) \leq d(f(x), f(y)) \quad x, y \text{ et } y \text{ sont}$$

dans la même classe

$$(1) \quad d(x, y) \geq d(f(x), f(y)) \quad x, y \text{ et } y \text{ sont}$$

diverses :

supposons que $[f(a), f(a), \dots, f(c_k)]$ soit une image uniformément mesurable de $[a, a, \dots, c_k]$ dans le sens



$a = (0, 1)$
 $b = (1, 0)$
 $c = (2.5, 0)$
 et prenons 4 fois le point c

considérons les trois points suivants que l'on désigne par a, b, c en deux classes

la procédure m' est pas bien structurée en 2 classes

images.

la procédure est donc bien adaptée par rapport aux
 est la partition qui minimise le critère.
 le qui est en contradiction avec le fait que $[a, a, \dots, c_k]$
 $\Delta(f(a), f(a), \dots, f(c_k)) < \Delta(a, a, \dots, c_k)$
 et donc

$$\begin{aligned}
 \sum_{x \in y} \|x - m_y\|^2 &= \sum_{x \in y} \|f(x) - f(m_y)\|^2 \\
 &> \sum_{x \in y} \|f(x) - f(m_y)\|^2 \\
 \sum_{x \in y} \|x - m_y\|^2 &= \sum_{x \in y} \|x - m_y\|^2
 \end{aligned}$$

Les données sont bien structurées en deux classes B_1 et

$$B_1 = \{a, b\}$$

$$B_2 = \{c\}$$

Parmi les partitions possibles de ces trois points en deux classes, choisissons celle qui minimise le critère Z

$$Z(\{a, b, \{c\}) = d^2(a, G(\{a, b\})) + d^2(b, G(\{a, b\})) = 1.36$$

$$Z(\{a\}, \{b, c\}) = 1.125$$

La procédure donne deux classes non bien structurées

$$A = \{a\}$$

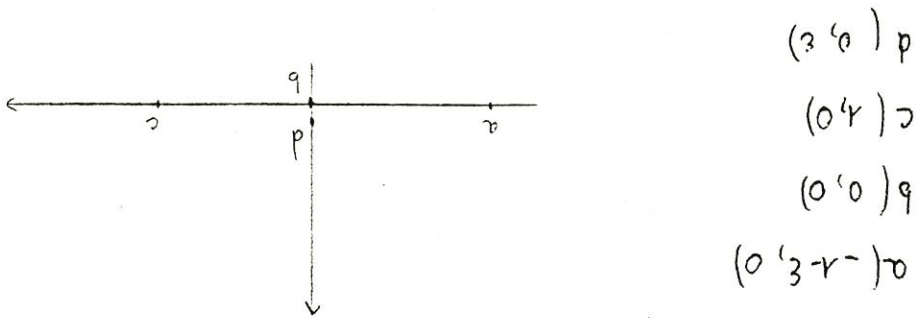
$$A = \{b, c\}$$

$$\text{en effet } diam(A) = 1.5 > d(a, b) = 1.2$$

La procédure n'est pas admissible par rapport aux

proportions des classes.

considérons les quatre points suivants



On désigne par G les points en deux classes qui minimisent la valeur du critère. La partition optimale est donnée par $G = \{a\}$, $G = \{b, c, d\}$. Par contre si on double tous les points de la classe G , le critère est optimisé par $G' = \{a, b, d\}$, $G' = \{c\}$.

Si on double tous les points d'une même classe on double tous les points d'une même classe. La partition optimale est donnée par $G = \{a\}$, $G = \{b, c, d\}$. La partition optimale est donnée par $G = \{a\}$, $G = \{b, c, d\}$. La partition optimale est donnée par $G = \{a\}$, $G = \{b, c, d\}$.

La procédure est admissible par rapport à ϕ .

non de classe

Il faut démontrer que si on enlève tous les points d'une même classe, soit ϕ , et que l'on applique la procédure aux points restants afin d'obtenir (R-1) classe, elle donne les mêmes classes à l'exception de ϕ .

Supposons que les échantillons donnés a', a'', \dots, a_{k-1}' soient
 différents de a, a', \dots, a_{k-1}
 cela voudrait dire que

$$L(a', a'', \dots, a_{k-1}') < L(a, a', \dots, a_{k-1})$$

et donc que

$$L(a', a'', \dots, a_{k-1}') + \sum_{x \in C_k} \|x - m_k\|^2 = L(a', \dots, a_{k-1}', c_k)$$

$$< L(a, a', \dots, a_{k-1}) + \sum_{x \in C_k} \|x - m_k\|^2 = L(a, a', \dots, a_{k-1}, c_k)$$

le qui n'est pas possible car (a, a', \dots, a_k)
 minimise le coût.

La procédure est donc admissible par rapport à
 l'union de classe.

La procédure n'est pas admissible par rapport
aux transformations monomorphes

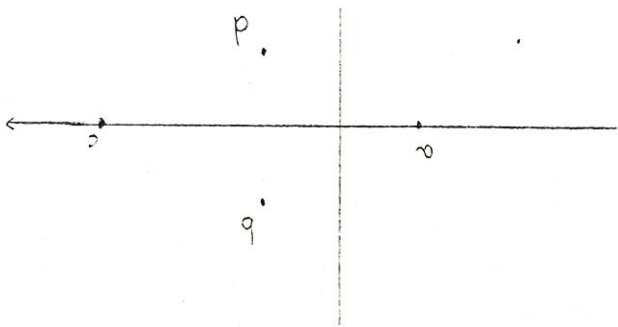
considérons les quatre points suivants

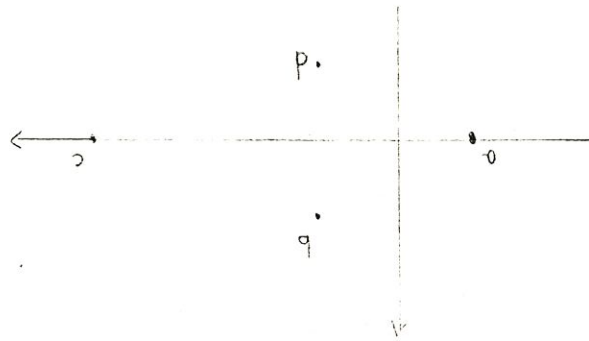
$$a = (-1, 0)$$

$$b = (1, 1)$$

$$c = (3, 1, 0)$$

$$d = (1, -1)$$





Remarquons que cela revient à transformer le point c.

	a	b	c	d
a	0	1.2	4.1	1.2
b		0	3.1	2
c			0	3.1
d				0

La solution optimale est $G = \{a\}$, $G = \{b, c, d\}$
 Appliquons une transformation monotone f aux éléments
 de la matrice des distances telle que $f(2.3) = 3.1$,
 et $f(4.1) = 5$. La matrice des distances devient

	a	b	c	d
a	0	1.2	4.1	1.2
b		0	1.3	2
c			0	2.3
d				0

La matrice des distances est
 doublons le point a